Concealed identity information detection with pupillometry in rapid serial visual presentation

Ivory Chen,[a] Aytaç Karabay,[a] Sebastiaan Mathôt,[a] Howard Bowman,[b,c] & Elkan G. Akyürek[a]

[a] Department of Experimental Psychology, University of Groningen, Groningen, the Netherlands

[b] School of Psychology, University of Birmingham, UK

[c] School of Computing, University of Kent, UK

Address correspondence to: Ivory Chen, Department of Experimental Psychology, University of Groningen, Grote Kruisstraat 2/1, 9712 TS, Groningen, The Netherlands. Phone: +31 (0)617482301. E-mail: y.n.chen@rug.nl

Abstract

The concealed information test (CIT) relies on bodily reactions to stimuli that are hidden in mind. However, people can use countermeasures, such as purposely focusing on irrelevant things, to confound the CIT. A new method designed to prevent the use of countermeasures, based on rapid serial visual presentation (RSVP), presents each stimulus on the fringe of awareness. Previous studies that used RSVP in combination with electroencephalography (EEG) showed that participants exhibit a clear reaction to their real first name, when they pretend to have a different name, even when they try to prevent such a reaction. Since EEG measures are not easily applicable outside the laboratory, we investigated here whether pupil size, which is easier to measure, is also able to detect concealed identity information. In our study, 31 participants were asked to adopt a fake name, and search for this name in an RSVP task, while their pupil sizes were recorded. Apart from this fake name, their real name and a control name also appeared in the task. We found that the pupil dilated more in response to the task-irrelevant real name, as compared to control names. However, while most participants showed this effect qualitatively, it was not statistically significant for most participants when analysed individually. Taken together, our results show that the current RSVP task with pupillometry can detect concealed identity information at a group level. Further development of the method is needed to create a valid and reliable concealed identity information detector at the individual level.

Keywords: Concealed information detection; RSVP; pupillometry

1 Introduction

Concealed information is something you hide from others in your mind. Concealed information can be anything, from your opinion on the clothing habits of your colleagues, to more serious crime-related information, for instance about a tool used in a crime, a particular date when a crime was or is going to be carried out, or identity information such as the name of a victim or an accomplice (Suchotzki & Gamer, 2018). To find reliable ways to detect such concealed crime-relevant information has long been a major goal of forensic scientists. With a reliable and valid method of detecting concealed information, more crimes could be solved and guilt as well as innocence could be more readily established.

The concealed information test (CIT) is a method that has been developed for this purpose. It has a high validity to detect concealed information, and it has been gradually improved in the past years (Ben-Shakhar & Elaad, 2003; Meijer et al., 2014; Verschuere & Kleinberg, 2016; Volz et al., 2018). The CIT was originally created by Lykken (1959) to test whether participants have crime-relevant knowledge. Generally, in the CIT, testers show participants crime-relevant stimuli (i.e., stimuli related to information that only the perpetrator has), and some neutral alternatives. The CIT rests on the assumption that implicit responses will be evoked by crime-relevant stimuli if participants already have that “guilty” knowledge. Thus, responses to the crime-relevant stimuli are compared to the responses to the neutral alternatives to assess whether the participant indeed has crime-relevant knowledge (Ben-Shakhar et al., 2011).

Several different methods of measuring responses in the CIT have been developed, most of which do not focus on overt behavior, but rather on autonomic responses. In the beginning, autonomic-nervous-system responses that indicate levels of arousal, such as heart rate, respiration and electrodermal activity, were measured (Kleiner, 2002; Rosenfeld et al., 2007). Additionally, in recent years, neuroimaging techniques, such as functional magnetic resonance imaging (fMRI), and stimulus-evoked brain potentials from electroencephalography (EEG), have become popular tools to record the reaction of the brain to the stimuli in the CIT (Gamer, 2014; Ganis, 2014; Hu et al., 2011; Mameli et al., 2010; Zeki et al., 2004). Although it might be noted that a measurement that is simpler and costs less, for example, electrodermal activity rather than fMRI, could be more easily adopted in forensic applications (Furedy, 2009), the CIT has proven capable of detecting concealed information across a range of measures (Ambach et al., 2010, 2019).

Nevertheless, despite extensive research efforts, the CIT cannot always reliably detect concealed information (Matsuda et al., 2012; Meijer et al., 2016). Its validity has been questioned because examinees can purposefully use physical and mental countermeasures to obscure the difference between responses to relevant and neutral alternatives (Ben-Shakhar, 2011; Peth et al., 2016). For example, they can bite their tongues to inflict pain or recall exciting memories when neutral alternatives are presented, and thereby confound the measurement (Mertens & Allen, 2008; Rosenfeld et al., 2004). The usefulness of any test that is only reliable with fully compliant examinees is obviously limited, and it is therefore important to find solutions to defeat countermeasures.

A new method, based on rapid serial visual presentation (RSVP), has recently been developed with the potential to eliminate this problem. In RSVP, a series of stimuli are presented sequentially in the same location with each stimulus visible only for about 100 milliseconds (Broadbent & Broadbent, 1987). This quick presentation on the fringe of awareness virtually eliminates the possibility of using countermeasures, because participants do not have enough time to exert top-down control over their responses to the stimuli. Bowman et al. (2013) first developed the RSVP paradigm for concealed identity information detection. The researchers measured EEG during a fake-name search task, and found that participants' real names triggered significant P3 potentials compared to control names, even when participants were explicitly instructed to hide their responses to their real name in various ways (Bowman et al., 2014). The results of Bowman and colleagues suggest that the RSVP paradigm reveals concealed identity information, is robust to countermeasures, and is therefore potentially more effective than the slower methods of presentation used to date.

Although the EEG results obtained by Bowman and colleagues (2013, 2014) with their RSVP task are promising, their method is not yet suitable for widespread application. Consider, for instance, that EEG facilities are not commonplace outside of university laboratories. Also with the CIT, it has been noted that new measures should be added to increase the probative force and the field use of CIT (Matsuda et al., 2012). It is thus important to assess whether the RSVP method is also effective with simpler measures that are more readily available in practice.

Some simpler measures related to eye responses, which can be acquired unobtrusively, have previously been shown to be effective dependent measurements in the CIT. For example, eye blinks after stimulus offset and fewer but longer fixations were found to indicate concealed information (Peth et al., 2016; Schwedes & Wentura, 2012; Leal & Vrij, 2010). In another study, involuntary inhibition of eye movements after stimulus onset was measured in a CIT with barely visible stimuli and verified as a valid indicator that discriminates 'terrorists' from 'innocents' (Rosenzweig & Bonneh, 2020). Eye movements, for which no sensors or electrodes need to be attached to a suspect, and for which only a sensitive video camera (i.e., an eye tracker) is needed, might thus be an efficient measurement also for RSVP-based detection of concealed information.

In this study, we focus on a particular kind of eye movement, namely the pupil response. Pupil responses are a promising measure in RSVP-based concealed information testing. First, pupil dilation and the P3 component of the event-related potential both reflect phasic responses in the locus coeruleus-norepinephrine (LC-NE) system. According to Nieuwenhuis et al.

(2011), a motivationally significant stimulus will evoke a dilation of the pupils as well as a P3. These two reactions are tightly linked to the activation of the LC-NE system (Koss, 1986; Murphy et al., 2011; Nieuwenhuis et al., 2005; Samuels & Szabadi, 2008). Processing task-relevant events will activate LC-NE's phasic response, followed by a pupil dilation and the P3 (Aston-Jones & Cohen, 2005). Since the P3 has proved to be an effective measure in RSVP-based CIT studies, it is reasonable to suppose pupil size will also be useful as a measure in the RSVP-based CIT.

Second, pupil dilation is capable of showing two different cognitive processes in the CIT. A CIT target requires participants to pay attention to and process task-related information. Pupil dilation reflects such effortful processes of cognitive control and attention required for responding to relevant information, while inhibiting irrelevant distractors (Cohen et al., 2015; Querino et al., 2015; Rondeel et al., 2015; van der Wel & van Steenbergen, 2018). Similarly, a critical concealed information stimulus, despite being task-irrelevant, will also attract attention and engage cognitive control processes. In CIT, it has been shown that the pupils also dilate when attention is allocated to new and salient stimuli that are task-irrelevant (Gilzenrat et al., 2010).

In addition, like other eye movements, pupil size has already been used in CIT studies and approved as an effective measure of concealed information. Lubow & Fein (1996) trained participants to be either guilty or innocent in a mock crime scenario, and then showed them some photographs of crime-relevant items (e.g., a green identification card or a face of a criminal), together with some crime-irrelevant items. 50-70% of the guilty participants and 100% of the innocent participants were correctly detected through the difference between pupil sizes to the crime-relevant and crime-irrelevant items. Another study also adopted pupil size as an indicator of concealed mock-crime-related knowledge (Seymour et al., 2013). With a hit rate of 83% and zero false-alarms, the authors were able to distinguish guilty from innocent participants with 92% accuracy.

Finally, even though the pupil response is slow, which at first glance might raise questions as to its effectiveness in the RSVP task, Wierda et al. (2012) showed that pupil dilation is able to reflect attention allocation and cognitive processing in RSVP. In their study, they used an attentional blink task of two target letters within a sequential stream of digits as distractors presented in an RSVP. Using a pupil dilation deconvolution method, the occurrence and timing of attentional processes, associated with the successive targets, were clearly tracked.

To sum up, these findings support the idea that pupil size could be an effective measure in RSVP-based CIT. The aim of the current study was thus to test the ability of the RSVP method in combination with a measure of pupil dilation to detect concealed identity information. For that purpose, we replicated the study by Bowman and colleagues, with some minor changes for pupil data recording, and used pupillometry instead of EEG. Each trial consisted of an RSVP stream, in which either a fake name, the participants' real name or a randomly selected control name appeared. The participants were asked to search for the fake name and to ignore their real name. Their pupil sizes were recorded during the task, and we tested whether their pupil responses to the real and control names differed reliably. The presence of such a difference would be indicative of the ability of the RSVP-pupil method to detect concealed identity information.

2 Method

2.1 Participants

Thirty-one participants took part in the experiment. All of them were first-year undergraduate students at the University of Groningen in the age group of 18-24 years (Mean: 19.29 years); there were 26 females and 5 males. All participants were native Dutch speakers. 26 participants were right-handed and 5 were left-handed. Participants had normal (uncorrected) vision. During the experiment, participants did not wear glasses, eye contacts or eye make-up. The study was conducted in accordance with the World Medical Association Declaration of Helsinki (2013) and approved by the ethical committee of the Psychology Department of the University of Groningen (approval number: PSY-18167-SP). Written informed consent was obtained prior to participation. Written and oral debriefing was provided after participation.

2.2 Apparatus and Stimuli

Participants were seated with their head on a chin rest with an adjustable height at a distance of approximate 60 cm from a 27'' LCD Iiyama PL2773H monitor with a display resolution of 1920×1080 pixels and a refresh rate of 100 Hz. On this monitor, stimuli were presented with OpenSesame 3.2.8 running on Windows 10 Enterprise. Pupil size was recorded in arbitrary units by an EyeLink 1000 (SR Research) during each trial using PyGaze (Dalmaijer et al., 2014).

We created a set of names for the experiment based on a database from the Meertens Institute for Dutch language and culture research (*Https://www.Meertens.Knaw.Nl/Nvb/Topnamen/Land/Nederland*). We first selected the 100 top Dutch names of each year from 1975 to 2014. Next, we excluded names consisting of more than 10 letters. This resulted in a set of 533 names with 281 female and 252 male names. From this name set, prior subsets of 15 possible names were selected randomly for each participant. A fake name and a control name were both selected from the unfamiliar names in the prior name subsets. Additionally, distractor names in each trial were selected pseudo-randomly from the set of 533 names: names with more than two identical consecutive letters were not allowed to be next to each other in one sequence. For example, ‘Dani’ and ‘Daniel’ had four identical consecutive letters; therefore, these two names could not be shown to participants directly after each other in one sequence. The distractor names were only used to form each name sequence and their presentation frequency was far less than the frequency of the fake, real and control names, which we were primarily concerned with.

We padded names on both sides with ‘+’ and ‘#’ characters so that the resulting string always consisted of 11 characters, as illustrated in Figure 1. Name stimuli were light grey (75% white; RGB: 190, 190, 190), 48 point, sans serif characters presented on a dark (RGB: 40, 40, 40) background. All the names were presented in the center of the screen. The visual angle for each name was 2.03° in height and from 8.88°-12.25° in width, with some variation because some letters are wider than others. Fixation dots were light grey (75% white; RGB: 190, 190, 190), and rendered in 48 point.

Figure 1. Examples of stimuli. List of example names used as stimuli. '+' and '#' were added to both sides of names in order to keep letters in the center positions.

2.3 Procedure

Prior to the start of the experiment, participants were presented with a subset of 15 possible female or male names matching their own gender from the name set. They were asked to indicate all the names of people they knew; these names were removed from the set of possible names to avoid confounds due to the familiarity of such names. After that, participants chose one of the remaining names as their target name for detection during the experiment; we refer to this as their "fake" name. If a participant removed all 15 names, a second round of 15 names would be shown until a fake name was selected. The primary task for the participants was to monitor the RSVP streams for the presence of this fake name. A single "control" name would also be selected from the remaining names after the fake name was selected, and their own "real" name was also added to the experiment. The pupil sizes of participants when they saw these three different critical names were used for later analyses.

Once a fake name and a control name were selected, the experiment started. As shown in Figure 2, each trial started with a drift-correction procedure (i.e., a one-point recalibration) followed by a fixation dot presented for 1000 milliseconds in order to establish a baseline pupil size. Then a stream of 11 names were displayed for 100 milliseconds each in a sequence. A dashed line (----------) or series of equal signs (=======) was presented for 100 milliseconds after the sequence. Participants were required to report this later as a secondary task that served to check whether their attention had remained on the stimulus presentation area throughout the stream. At the end of each trial, a fixation dot was shown again for 2000 milliseconds, to allow capture of the full pupil response, which continues for some time after stimulus presentation.

Participants were asked to try not to blink from the appearance of the first fixation dot to the disappearance of the second fixation dot.

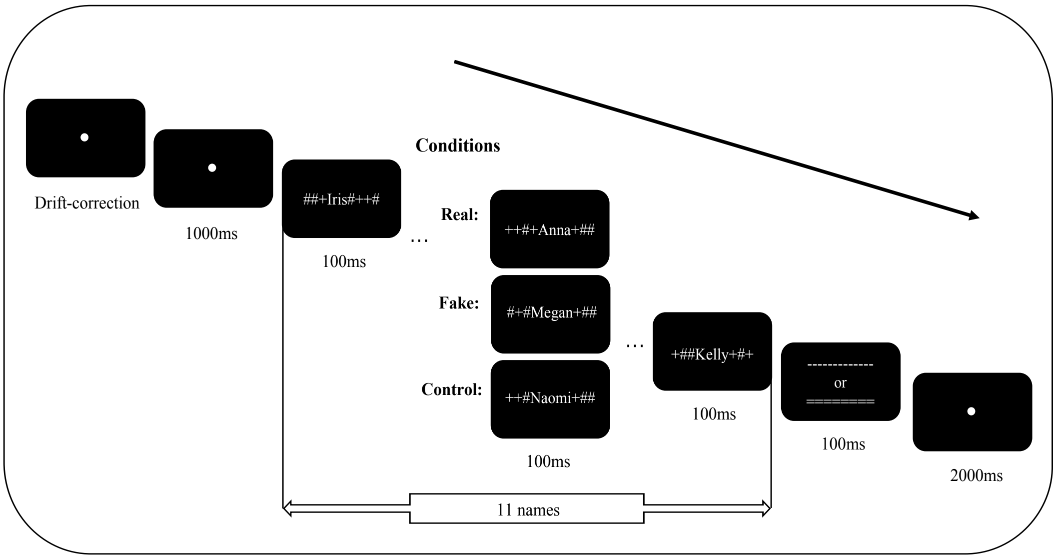

Figure 2. Trial sequence. Each trial started with a drift-correction procedure. Next, a fixation dot was displayed for another 1000ms to enable pupil size to return to baseline. This was followed by a sequence of 11 names, each presented for 100ms. 10 of these 11 names were distractors. At a random position between 5 and 9, one name defined three conditions. In the figure, Anna is the participant's *real* name; Megan is the *fake* name chosen by the participant; and Naomi was selected as the *irrelevant* (control) name, from a set of 15 random, unfamiliar names, presented to the participant before the trials. After the name sequence, a dashed line or equal signs appeared for 100ms to keep participants' attention fixed throughout the entire stream. At the end, a 2000ms fixation dot was shown to allow the pupil reaction to develop fully.

In the sequence, 10 of 11 names were distractors. One critical name, at a random position between 5 (earliest) and 9 (latest) in the sequence, defined three different conditions: this name was the participants' real name (to which no behavioral response should be made), the fake name (the name that participants had selected previously and were instructed to respond to), or a control name (a name randomly selected from unfamiliar names of the previous sets with 15 names each, serving as a baseline that was matched in presentation frequency to the real and fake names, and which also did not require a response). Each condition was represented in 60 out of 180 trials separately.

At the end of each trial, participants were asked to answer two questions: 1. Did you see your (fake) name? 2. Did you see ---------- or =======? Participants answered the two questions by pressing the ‘F’ or ‘J’ keys on a standard QWERTY keyboard. Whether ‘F’ was for ‘Yes’ or ‘No’ was counterbalanced based on participant number parity to balance the response mapping between participants.

There were two sessions in the whole experiment: a practice session and an experimental session. The practice session consisted of 20 trials, during which participants received trial-based feedback (a smiley face for correct answers and a frowney face for wrong answers) to indicate whether they had responded correctly or not to the first and second question. The experimental session consisted of 180 trials divided into 6 blocks with a break between each pair of blocks. Conditions (real, fake, control) were randomly mixed within blocks. All participants took part in all three conditions, and the practice session, and thus completed 200 trials in total.

2.4 Data processing and analysis

All data and the analysis scripts are publicly accessible on the Open Science Framework (*Https://osf.io/9fkpm/*).

We first analyzed how accurately participants responded to question 1 (did you see your (fake) name?) and question 2 (did you see ---------- or =======?). The response accuracy for question 1 reflects how well participants were able to detect their fake names, and therefore whether the difficulty of the RSVP task was reasonable in the sense that it did not show a floor or ceiling effect. The accuracy for question 2 reflects to what extent participants maintained their attention on the RSVP stream throughout the trial, also after the appearance of the critical name.

On each trial, 800 samples of pupil-size data were obtained, given that the EyeLink 1000 sampled at 250 Hz and a whole RSVP trial was 3200ms from the start of a name sequence to the end of the following fixation. We down-sampled the signal to 25 Hz, leaving 80 samples per trial for each participant. Then we baselined the pupil sizes by subtracting the mean of the first three samples in the pupil traces for each trial separately. After baselining, we locked the pupil sizes to the onsets of the critical name positions, such that timepoint 0 corresponded to the onset of the name regardless of at which position in the stream the name was presented. Finally, because of the initial pupil constriction that occurs generically after the onset of stimuli (Mathôt, 2018), we deleted the first 200ms of data, leaving 2200 ms (55 samples) for each participant per trial.

We ran a sample-by-sample linear mixed effects analysis on the group level to check for possible effects on pupil size of the fake and real names, compared to the control names. More specifically, for each 40 ms sample separately we conducted a linear mixed effects analysis with baseline-corrected pupil size as dependent measure, condition as fixed effect (with control as reference value), and by-participant random intercepts and slopes, using the lme4 and lmerTest packages for R. For sample-by-sample analyses, we considered an effect reliable if $p < .05$ for at least 200 ms (5 samples).

Our predictions were twofold. First, if pupil size in the fake condition is significantly larger than in the control condition, then this indicates that the fake names (the task-relevant stimuli) elicited a reaction, and that this (presumably attentional) reaction can be detected by pupillometry in RSVP at the group level, as would be expected from prior research. Second, if pupil size in the real condition is significantly larger than in the control condition, then this indicates that the real name also elicited a detectable reaction, despite this name being task-irrelevant. This second effect is what we were primarily interested in, because it would indicate that pupil size is useful as a tool for the RSVP-based concealed-information detection.

In addition, we tested whether pupil size was modulated by learning, fatigue, or habituation over the course of the experimental session. Specifically, we wanted to test whether the difference between the real and control conditions became smaller over time, as participants might have learned to 'desensitize' to the presentation of their own name, for instance. To test this, we divided trials into two sets: the first 90 trials and the last 90 trials of each participant. Then we tested these two sets separately using the same sample-by-sample linear mixed effects analysis as described above to check the effect of the critical names. In addition, in a separate sample-by-sample analysis similar to the one described above, we included both trial number and condition as fixed effects (and corresponding by-participant random effects). This analysis allowed us to check whether trial number (testing time) robustly modulates the differences between the three critical name conditions.

To further check whether our RSVP task in combination with pupillometry is useful as a tool for concealed-information detection, we wanted to establish whether the effects that were measured at the group level (see above) could also be measured reliably at the individual level. To do this, we conducted a leave-one-out analysis. For each participant separately, we first determined the peak indices (time points) at which the pupil-size difference between the real and control condition, as well as the fake and control condition, was largest for all other 30 participants. Then we conducted t-tests separately for the pupil-size difference between the real and control condition, as well as between the fake and control condition, at these corresponding peak indices. The general logic behind this approach is that we use the data from all-but-one participants as a 'temporal localizer' to determine the optimal time point to test for a given participant. If pupil size for the fake condition is significantly different from the control condition for each participant, then this would mean that our approach is able to detect the (attentional) reaction to task-relevant target names, even for individual participants. Similarly, if pupil size for the real condition is significantly different from the control condition for each participant, then this would mean that our approach is able to detect the presence of concealed information, even for individual participants. Given the relatively small effect sizes that are commonly observed in psychological research, and given that neither our task parameters nor analysis approach have yet been explored systematically in the present context, this is a lot to ask of our data. Nevertheless, with an eye towards practical applications, the individual-participant reliability is important to assess.

3 Results

3.1 Task performance

Participants responded to questions 1 and 2 with an accuracy of 89% and 95.4% respectively. Thus, participants were well able to detect their fake names in the RSVP sequences, and to keep their attention on the stimuli throughout each trial.

3.2 Pupil data

For the group-level analysis, the sample-by-sample linear mixed effects analysis showed that pupil size in the fake condition was significantly larger than in the control condition from about 360ms to 2200ms. The pupil size in the real condition was significantly larger than in the control condition from about 320ms to 1120ms. The results are visualised in Figure 3.

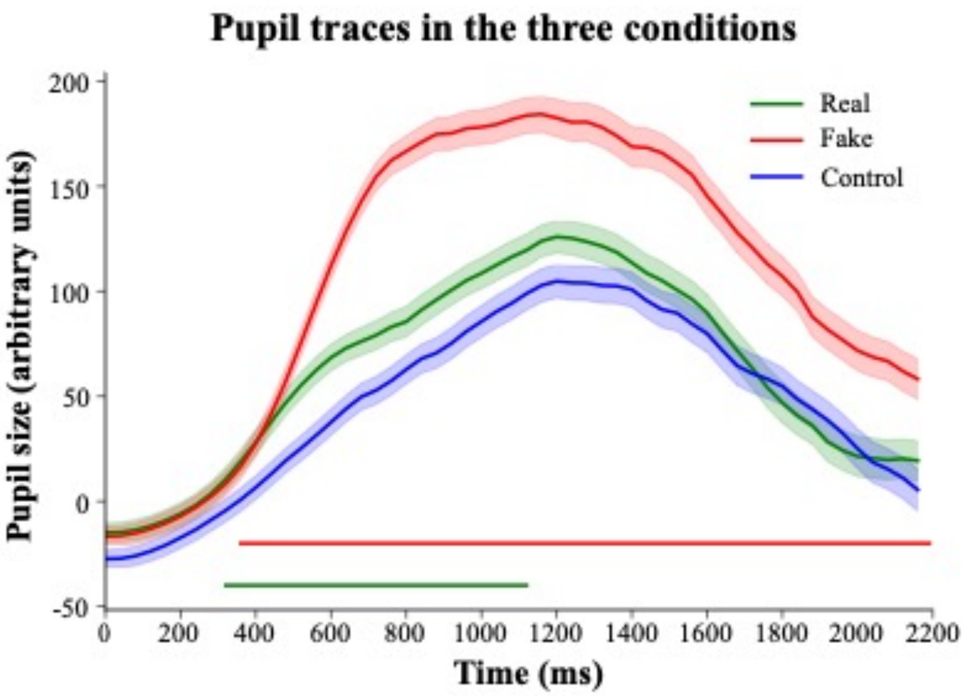

Figure 3. Pupil traces in three conditions. Pupil size average (N=31) is shown over time, in the fake condition (response to the assigned, task-relevant fake name; red line), the real condition (response to the participant's real name; green line), and the control condition (irrelevant name; blue line). The x-axis indicates time in milliseconds where 0 corresponds to 200ms after the onset of the name; the y-axis indicates baseline-corrected pupil size in arbitrary units. The red straight line shows the timepoints (from about 360ms to 2200ms) when the pupil size in the fake condition was significantly larger than in the control condition. The green straight line shows the timepoints (from about 320ms to 1120ms) when the pupil size in the real condition was significantly larger than in the control condition.

Looking at the learning effect, in the first 90 trials, the sample-by-sample linear mixed effects analysis showed that pupil size in the fake condition was significantly larger than in the control condition from about 360ms to 2200ms. Pupil size in the real condition was significantly larger than in the control condition from about 400ms to 1080ms and from about 1120ms to 1360ms. In the second 90 trials, only in the fake condition, was pupil size statistically significantly larger than in the control condition, which was from about 440ms to 2200ms but there was no longer a reliable difference between the real and control conditions. These results are shown in Figure 4.

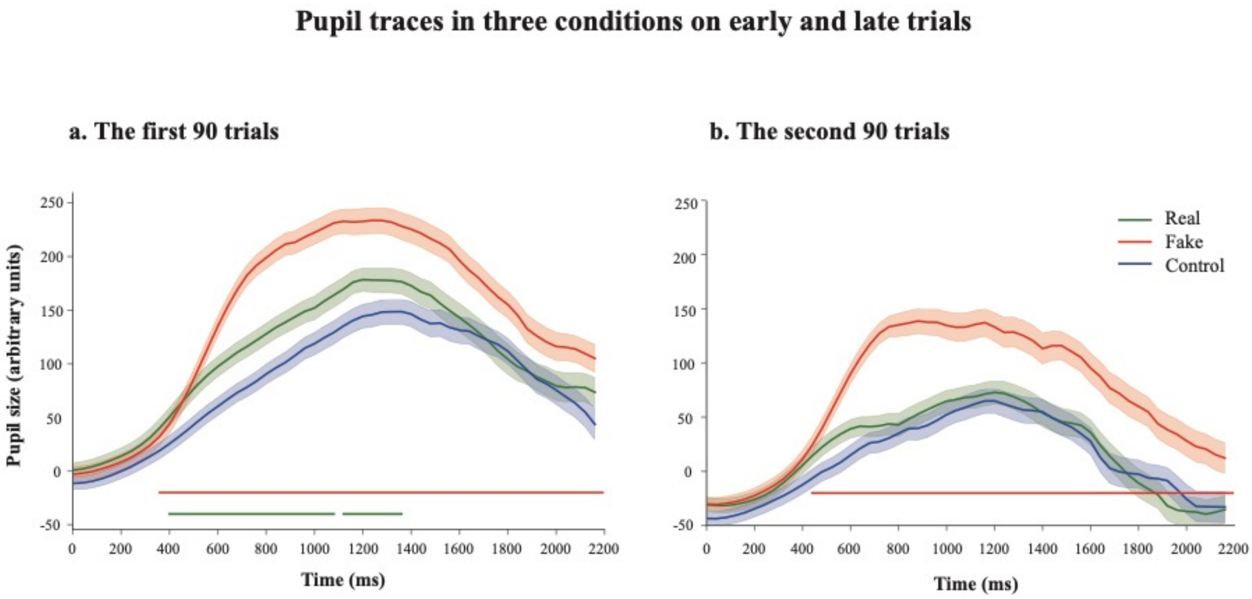

Figure 4. Pupil traces in three conditions on early (a) and late (b) trials. Pupil size average (N=31) is shown over time, in the fake condition (response to the assigned, task-relevant fake name; red line), the real condition (response to the participant's real name; green line), and the control condition (irrelevant name; blue line) in the first (a) and second (b) half of the experiment. The x-axis indicates time in milliseconds, where 0 corresponds to 200ms after the onset of the name; the y-axis indicates baseline-corrected pupil size in arbitrary units. Panel a) shows the results for the first 90 trials. The red straight line shows the duration (from about 360ms to 2200ms) when pupil size in the fake condition is significantly larger than in the control condition. The green straight line shows the duration (from about 400ms to 1080ms and from about 1120ms to 1360ms) when pupil size in the real condition is significantly larger than in the control condition. Panel b) shows the second 90 trials. The red straight line shows the duration (from about 440ms to 2200ms) when pupil size in the fake condition is significantly larger than in the control condition. Pupil sizes in the real and control conditions are not significantly different.

However, when we tested the learning effect more rigorously as an interaction between trial number and condition, we did not find a reliable interaction at any point in time. We only found that pupil size decreased over time, possibly as an effect of increasing fatigue. Taken together, the results suggest that pupil size is sensitive to concealed identity information from the very start of the testing session; if anything, this sensitivity seems to decrease over time, which suggests that there may not be much to gain from very long testing sessions.

In the individual-level leave one out analyses, although the vast majority (23 of 31) of participants showed a qualitative effect in the predicted direction, only six participants out of 31 showed a significant pupil-size difference between the real and control conditions, with a mean p-value of 0.214, and a median p-value of 0.217. In contrast, 21 participants out of 31 showed a significant pupil-size difference between the fake and control conditions, with a mean p-value of 0.058, and a median p-value of 0.013. These results are shown in Figure 5.

Individual effect sizes

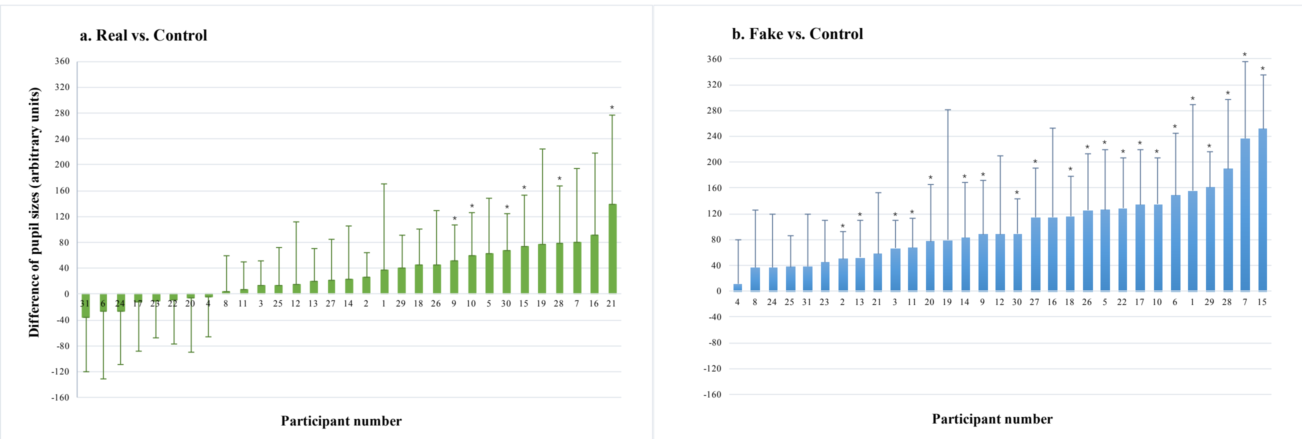

Figure 5. Individual effect sizes. The difference between pupil sizes for the real and control conditions of each participant (N=31) is shown in panel a, (sorted by effect size). Six participants showed a significant difference between pupil size for real and control (marked by *). The difference between pupil size for fake and control of each participant (N=31) is shown in panel b (again sorted by effect size). Differences between pupil sizes for fake and control of 21 participants are statistically significant (marked by *) The error bars indicate the 95% confidence interval of each difference per participant.

4. Discussion

As shown empirically in this study, pupil size in a rapid serial visual presentation (RSVP) task can provide valuable information regarding concealed identity information. More specifically, we observed pupil dilation in response to participants' real names, even though participants were instructed to respond only to a 'fake' name that they had selected at the start of the experiment.

This finding is similar to the results Bowman et al. (2014) found in their EEG study. In their study, participants' real names triggered a significant P3 component of the event-related potential, compared to control names at the group level, while they tried to search for 'fake names' and ignore 'real names'. In this study, we similarly found that concealed identity information evoked a detectable pupillary response. Thus, as a cheaper, more available, less technically complex, and non-invasive measure, pupillometry is effective in RSVP for concealed identity information detection on the group level. Thereby, pupillometry can be considered as a potentially promising measure to be used widely in practice with a RSVP concealed information detector. This result also corresponds to the previous findings that the P3 component and pupil dilation both reflect the activation of phasic responses in the locus coeruleus-norepinephrine (LC-NE) system (Koss, 1986; Murphy et al., 2011; Nieuwenhuis et al., 2005, 2011; Samuels & Szabadi, 2008). Thus, it seems fair to assume that target-triggered cognitive processes and attention allocation caused by the real names activated the LC-NE system, which is followed by (the P3 component and) dilated pupil size.

We furthermore observed a second practical advantage in our study. The real-name effect on pupil size was found already at the start of the experiment and it only seemed to decline over time. Even though no evidence for a strong time on task effect was found, it suggests that we might be able to avoid extensive measuring sessions, which may actually reduce the effect of interest. To further benefit from the early onset of the present difference between real and control names, it may be advisable to keep real names only in the testing sessions and not include it in the practice trials.

However, the current outcomes also showed that while most participants showed the desired effect qualitatively, it was not statistically significant for most participants when analyzed individually. In the RSVP-EEG study of Bowman and colleagues (2013), the authors were able to increase the individual success rate considerably by using Fisher's method to combine data from three electrodes (Fz, Cz and Pz). They succeeded in obtaining a significant p-value for the difference between the real and the control conditions for each of their 15 participants. Pupil data does not immediately offer a similar option, since there is just a single signal being measured. It might nevertheless be an avenue for future research to attempt to combine a pupil measure with a secondary measure. Measures of (micro-) saccades, fixation duration, or even other physiological indicators, such as heart rate, may present suitable candidates.

Modifying the current design for increased power at the individual level is another approach that might be further explored. Such modifications may include adjusting the presentation rate and other properties of the stimuli used in the task. Based on the current design and result, we may also suggest that future research could consider choosing a-priori time windows to look at during data analysis rather than focusing on a time-point. Since most of the participants showed a qualitative effect in the predicted direction, a more sensitive analytic method may reflect the real-name effect on pupil size better. For instance, 500ms to 1600ms after the onset of the critical names (corresponding to 300 to 1400 ms in Figure 3 and 4, after subtracting the first 200 ms of deleted data) would be a sensitive time window to assess possible differences in pupil size.

In conclusion, we have shown that pupil size is sensitive to the presence of concealed identity information (the participants' own name) in an RSVP task. This implies that pupil size is a promising measure for detecting concealed information. However, further refinement is required to improve the task's sensitivity so that concealed identity information can be detected reliably even at the level of individual participants, and we have offered several suggestions for follow-up. In summary, the present study is a promising proof of concept for using pupillometry as a measure of concealed information detection.

References

Ambach, W., Assmann, B., Wielandt, B., & Vaitl, D. (2019). Evaluative observation in a concealed information test. *Frontiers in Psychiatry*, *10*, 636. https://www-doi-org.proxy-ub.rug.nl/10.3389/fpsyt.2019.00636

Ambach, W., Bursch, S., Stark, R., & Vaitl, D. (2010). A concealed information test with multimodal measurement. *International Journal of Psychophysiology*, *75*(3), 258–267. https://www-doi-org.proxy-ub.rug.nl/10.1016/j.ijpsycho.2009.12.007

Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: Adaptive Gain and Optimal Performance. *Annual Review of Neuroscience*, *28*(1), 403–450. https://www-doi-org.proxy-ub.rug.nl/10.1146/annurev.neuro.28.061604.135709

Ben-Shakhar, G. (2011). Countermeasures. In B. Verschuere, G. Ben-Shakhar, & E. Meijer (Eds.), *Memory Detection: Theory and Application of the Concealed Information Test* (pp. 200–214). Cambridge University Press.

Ben-Shakhar, G., & Elaad, E. (2003). The validity of psychophysiological detection of information with the guilty knowledge test: A meta-analytic review. *Journal of Applied Psychology*, *88*(1), 131–151. https://www-doi-org.proxy-ub.rug.nl/10.1037/0021-9010.88.1.131

Ben-Shakhar, G., Meijer, E., & Verschuere, B. (2011). *Memory detection: Theory and application of the concealed information test*. Cambridge University Press.

Bowman, H., Filetti, M., Alsufyani, A., Janssen, D., & Su, L. (2014). Countering countermeasures: detecting identity lies by detecting conscious breakthrough. *PLoS ONE*, *9*(3), e90595. https://www-doi-org.proxy-ub.rug.nl/10.1371/journal.pone.0090595

Bowman, H., Filetti, M., Janssen, D., Su, L., Alsufyani, A., & Wyble, B. (2013). Subliminal salience search illustrated: EEG identity and deception detection on the fringe of awareness. *PLoS ONE*, *8*(1), e54258. https://www-doi-org.proxy-ub.rug.nl/10.1371/journal.pone.0054258

Broadbent, D. E., & Broadbent, M. H. P. (1987). From detection to identification: Response to multiple targets in rapid serial visual presentation. *Perception & Psychophysics*, *42*(2), 105–113. https://www-doi-org.proxy-ub.rug.nl/10.3758/BF03210498

Cohen, N., Moyal, N., & Henik, A. (2015). Executive control suppresses pupillary responses to aversive stimuli. *Biological Psychology*, *112*, 1–11. https://www-doi-org.proxy-ub.rug.nl/10.1016/j.biopsycho.2015.09.006

Dalmaijer, E. S., Mathôt, S., & Van der Stigchel, S. (2014). PyGaze: An open-source, cross-platform toolbox for minimal-effort programming of eyetracking experiments. *Behavior Research Methods*, *46*(4), 913–921. https://www-doi-org.proxy-ub.rug.nl/10.3758/s13428-013-0422-2

Furedy, J. J. (2009). The concealed information test as an instrument of applied differential psychophysiology: Methodological considerations. *Applied Psychophysiology and Biofeedback*, *34*(3), 149–160. https://www-doi-org.proxy-ub.rug.nl/10.1007/s10484-009-9097-y

Gamer, M. (2014). Mind reading using neuroimaging: Is this the future of deception detection? *European Psychologist*, *19*(3), 172–183. https://www-doi-org.proxy-ub.rug.nl/10.1027/1016-9040/a000193

Ganis, G. (2014). Deception detection using neuroimaging. In P. A. Granhag, A. Vrij, & B. Verschuere (Eds.), *Detecting Deception* (pp. 105–121). John Wiley & Sons, Ltd. https://www-doi-org.proxy-ub.rug.nl/10.1002/9781118510001.ch5

Gilzenrat, M. S., Nieuwenhuis, S., Jepma, M., & Cohen, J. D. (2010). Pupil diameter tracks changes in control state predicted by the adaptive gain theory of locus coeruleus function. *Cognitive, Affective, & Behavioral Neuroscience*, *10*(2), 252–269. https://www-doi-org.proxy-ub.rug.nl/10.3758/CABN.10.2.252

Hu, X., Wu, H., & Fu, G. (2011). Temporal course of executive control when lying about self- and other-referential information: An ERP study. *Brain Research*, *1369*, 149–157. https://www-doi-org.proxy-ub.rug.nl/10.1016/j.brainres.2010.10.106

Kleiner, M. (2002). *Handbook of polygraph testing*. Academic Press.

Koss, M. C. (1986). Pupillary dilation as an index of central nervous system α2-adrenoceptor activation. *Journal of Pharmacological Methods*, *15*(1), 1–19. https://www-doi-org.proxy-ub.rug.nl/10.1016/0160-5402(86)90002-1

Lubow, R. E., & Fein, O. (1996). Pupillary size in response to a visual guilty knowledge test: New technique for the detection of deception. *Journal of Experimental Psychology: Applied*, *2*(2), 164–177. https://www-doi-org.proxy-ub.rug.nl/10.1037/1076-898X.2.2.164

Lykken, D. T. (1959). The GSR in the detection of guilt. *Journal of Applied Psychology*, *43*(6), 385–388. https://www-doi-org.proxy-ub.rug.nl/10.1037/h0046060

Mameli, F., Mrakic-Sposta, S., Vergari, M., Fumagalli, M., Macis, M., Ferrucci, R., Nordio, F., Consonni, D., Sartori, G., & Priori, A. (2010). Dorsolateral prefrontal cortex specifically processes general – but not personal – knowledge deception: Multiple brain networks for lying. *Behavioural Brain Research*, *211*(2), 164–168. https://www-doi-org.proxy-ub.rug.nl/10.1016/j.bbr.2010.03.024

Mathôt, S. (2018). Pupillometry: Psychology, physiology, and function. *Journal of Cognition*, *1*(1), 16. https://www-doi-org.proxy-ub.rug.nl/10.5334/joc.18

Matsuda, I., Nittono, H., & Allen, J. J. B. (2012). The current and future status of the concealed information test for field use. *Frontiers in Psychology*, *3*. https://www-doi-org.proxy-ub.rug.nl/10.3389/fpsyg.2012.00532

Meijer, E. H., Selle, N. K., Elber, L., & Ben-Shakhar, G. (2014). Memory detection with the concealed information test: A meta analysis of skin conductance, respiration, heart rate, and P300 data: CIT meta-analysis of SCR, respiration, HR, and P300. *Psychophysiology*, *51*(9), 879–904. https://www-doi-org.proxy-ub.rug.nl/10.1111/psyp.12239

Meijer, E. H., Verschuere, B., Gamer, M., Merckelbach, H., & Ben-Shakhar, G. (2016). Deception detection with behavioral, autonomic, and neural measures: Conceptual and methodological considerations that warrant modesty: Deception research: Methodological considerations. *Psychophysiology*, *53*(5), 593–604. https://www-doi-org.proxy-ub.rug.nl/10.1111/psyp.12609

Mertens, R., & Allen, J. J. B. (2008). The role of psychophysiology in forensic assessments: Deception detection, ERPs, and virtual reality mock crime scenarios. *Psychophysiology*, *45*(2), 286–298. https://www-doi-org.proxy-ub.rug.nl/10.1111/j.1469-8986.2007.00615.x

Murphy, P. R., Robertson, I. H., Balsters, J. H., & O'connell, R. G. (2011). Pupillometry and P3 index the locus coeruleus-noradrenergic arousal function in humans: Indirect markers of locus coeruleus activity. *Psychophysiology*, *48*(11), 1532–1543. https://www-doi-org.proxy-ub.rug.nl/10.1111/j.1469-8986.2011.01226.x

Nieuwenhuis, S., Aston-Jones, G., & Cohen, J. D. (2005). Decision making, the P3, and the locus coeruleus-norepinephrine system. *Psychological Bulletin*, *131*(4), 510–532. https://www-doi-org.proxy-ub.rug.nl/10.1037/0033-2909.131.4.510

Nieuwenhuis, S., De Geus, E. J., & Aston-Jones, G. (2011). The anatomical and functional relationship between the P3 and autonomic components of the orienting response: P3 and orienting response. *Psychophysiology*, *48*(2), 162–175. https://www-doi-org.proxy-ub.rug.nl/10.1111/j.1469-8986.2010.01057.x

Peth, J., Suchotzki, K., & Gamer, M. (2016). Influence of countermeasures on the validity of the concealed information test: Influence of countermeasures on CIT validity. *Psychophysiology*, *53*(9), 1429–1440. https://www-doi-org.proxy-ub.rug.nl/10.1111/psyp.12690

Querino, E., dos Santos, L., Ginani, G., Nicolau, E., Miranda, D., Romano-Silva, M., & Malloy-Diniz, L. (2015). Cognitive effort and pupil dilation in controlled and automatic processes. *Translational Neuroscience*, *6*(1), 168–173. https://www-doi-org.proxy-ub.rug.nl/10.1515/tnsci-2015-0017

Rondeel, E. W. M., van Steenbergen, H., Holland, R. W., & van Knippenberg, A. (2015). A closer look at cognitive control: Differences in resource allocation during updating, inhibition and switching as revealed by pupillometry. *Frontiers in Human Neuroscience*, *9*. https://www-doi-org.proxy-ub.rug.nl/10.3389/fnhum.2015.00494

Rosenfeld, J.P., Shue, E., & Singer, E. (2007). Single versus multiple probe blocks of P300-based concealed information tests for self-referring versus incidentally obtained information. *Biological Psychology*, *74*(3), 396–404. https://www-doi-org.proxy-ub.rug.nl/10.1016/j.biopsycho.2006.10.002

Rosenfeld, J.Peter, Soskins, M., Bosh, G., & Ryan, A. (2004). Simple, effective countermeasures to P300-based tests of detection of concealed information. *Psychophysiology*, *41*(2), 205–219. https://www-doi-org.proxy-ub.rug.nl/10.1111/j.1469-8986.2004.00158.x

Rosenzweig, G., & Bonneh, Y. S. (2020). Concealed information revealed by involuntary eye movements on the fringe of awareness in a mock terror experiment. *Scientific Reports*, *10*(1), 14355. https://www-doi-org.proxy-ub.rug.nl/10.1038/s41598-020-71487-9

Samuels, E., & Szabadi, E. (2008). Functional neuroanatomy of the noradrenergic locus coeruleus: Its roles in the regulation of arousal and autonomic function part I: Principles of functional organisation. *Current Neuropharmacology*, *6*(3), 235–253. https://www-doi-org.proxy-ub.rug.nl/10.2174/157015908785777229

Schwedes, C., & Wentura, D. (2012). The revealing glance: Eye gaze behavior to concealed information. *Memory & Cognition*, *40*(4), 642–651. https://www-doi-org.proxy-ub.rug.nl/10.3758/s13421-011-0173-1

Seymour, T. L., Baker, C. A., & Gaunt, J. T. (2013). Combining blink, pupil, and response time measures in a concealed knowledge test. *Frontiers in Psychology*, *3*. https://www-doi-org.proxy-ub.rug.nl/10.3389/fpsyg.2012.00614

Sharon Leal & Aldert Vrij. (2010). The occurrence of eye blinks during a guilty knowledge test. *Psychology, Crime & Law*, *16*, 349–357. https://www-doi-org.proxy-ub.rug.nl/10.1080/10683160902776843

Suchotzki, K., & Gamer, M. (2018). Alcohol facilitates detection of concealed identity information. *Scientific Reports*, *8*(1), 7825. https://www-doi-org.proxy-ub.rug.nl/10.1038/s41598-018-25811-z

van der Wel, P., & van Steenbergen, H. (2018). Pupil dilation as an index of effort in cognitive control tasks: A review. *Psychonomic Bulletin & Review*, *25*(6), 2005–2015. https://www-doi-org.proxy-ub.rug.nl/10.3758/s13423-018-1432-y

Verschuere, B., & Kleinberg, B. (2016). ID-Check: Online concealed information test reveals true identity. *Journal of Forensic Sciences*, *61*, S237–S240. https://www-doi-org.proxy-ub.rug.nl/10.1111/1556-4029.12960

Volz, K., Bahr, L., Heinrichs, M., Vaitl, D., & Ambach, W. (2018). The concealed information test is susceptible to misleading information. *Journal of Forensic Sciences*, *63*(5), 1419–1428. https://www-doi-org.proxy-ub.rug.nl/10.1111/1556-4029.13718

Wierda, S. M., van Rijn, H., Taatgen, N. A., & Martens, S. (2012). Pupil dilation deconvolution reveals the dynamics of attention at high temporal resolution. *Proceedings of the National Academy of Sciences*, *109*(22), 8456–8460. https://www-doi-org.proxy-ub.rug.nl/10.1073/pnas.1201858109

World Medical Association Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects. (2013). *JAMA*, *310*(20), 2191. https://www-doi-org.proxy-ub.rug.nl/10.1001/jama.2013.281053

Zeki, S., Goodenough, O. R., Spence, S. A., Hunter, M. D., Farrow, T. F. D., Green, R. D., Leung, D. H., Hughes, C. J., & Ganesan, V. (2004). A cognitive neurobiological account of deception: Evidence from functional neuroimaging. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *359*(1451), 1755–1762. https://www-doi-org.proxy-ub.rug.nl/10.1098/rstb.2004.1555

CRediT author statement

Elkan G. Akyürek: Conceptualization (lead); formal analysis (supporting); methodology (supporting); supervision (lead); visualization (supporting); writing - original draft preparation (supporting); write - review & editing (lead); funding acquisition (supporting).

Howard Bowman: Conceptualization (supporting); formal analysis (supporting); write - review & editing (supporting).

Ivory Chen: Data curation (lead); formal analysis (supporting); investigation (lead); methodology (supporting); project administration (lead); visualization (lead); writing - original draft preparation (lead); write - review & editing (supporting); funding acquisition (lead).

Aytaç Karabay: Conceptualization (supporting); methodology (lead); Data curation (supporting); software (lead); write - review & editing (supporting).

Sebastiaan Mathôt: Data curation (supporting); formal analysis (lead); software (supporting); supervision (supporting); visualization (supporting); writing - original draft preparation (supporting); writing - review & editing (supporting); funding acquisition (supporting).

Author Notes

I. Chen was supported by the Chinese Scholarship Council, grant 202007720091.

Please direct author reprint requests to: Ivory Chen at y.n.chen@rug.nl