# Stereotypical behavioural cues — but not their order — influence credibility judgements

Received: 11 December 2018 | Revised: 17 October 2019 | Accepted: 5 February 2020

DOI: 10.1002/jip.1543

RESEARCH ARTICLE

WILEY

# Stereotypical behavioural cues — but not their order — influence credibility judgements

Glynis Bogaard | Ewout H. Meijer

Department of Clinical Psychological Science, Section Forensic Psychology, Maastricht University, Maastricht, The Netherlands

**Correspondence**
Glynis Bogaard, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands.
Email: glynis.bogaard@maastrichtuniversity.nl

## Abstract

To what extent stereotypical deceptive behaviours such as gaze aversion and fidgeting actually influence people's credibility judgements remain largely unknown. In this study, we directly manipulated the presence/absence of such behaviours to investigate this. Participants were shown four truthful videos in which we manipulated the presence of stereotypical cues and asked them to judge how credible the person in each video is. Moreover, research consistently shows that decision making is influenced by various cognitive biases. One example is the primacy effect, which implies that people form an opinion early in the decision process. Information acquired early will have the largest influence on how subsequent information will be interpreted. To investigate a possible primacy effect, we also manipulated whether these cues were present towards the beginning or the end of the video (i.e. the timing of the manipulation). In line with our expectations, the presence of stereotypical cues significantly lowered the observed credibility, showing that the presence of these cues indeed influences credibility judgements. The timing of the cues had no effect.

KEYWORDS

credibility assessment, cues to deception, nonverbal behaviour

## 1 | INTRODUCTION

Both lay people and professionals report to rely heavily on nonverbal cues such as gaze aversion when making judgements about someone's credibility (Bogaard, Meijer, Vrij, & Merckelbach, 2016; Masip & Herrero, 2015; The Global

Deception Research Team, 2006). Meta-analytic research (DePaulo et al., 2003), however, revealed that very few nonverbal cues correlate with lying, and even if a correlation was found it was typically small. Gaze aversion, for example, showed no relationship with deception ($d$ = 0.03). Not surprisingly, a number of authors have argued that this reliance on invalid cues explains the poor deception detection accuracy people typically display (e.g. Strömwall, Granhag, & Hartwig, 2004).

Recent meta-analytical findings have contested the view that poor credibility judgement performance can be explained by incorrect beliefs about what cues signal deception. Hartwig and Bond (2011; see also Hartwig and Granhag, 2014) showed large discrepancies between the cues people report when they are asked how they detect lies, and what they actually base their credibility judgement on. Using a lens model approach, their meta-analytic research showed that stereotypical nonverbal cues (e.g. gaze aversion, fidgeting) contributed only little to credibility judgements. Instead, these judgements were based on impressions of incompetence/ambivalence, and lack of spontaneity.

Yet, several empirical studies hint at a role of nonverbal cues in credibility judgements. For example, early research has shown that people displaying weird nonverbal behaviour (i.e. raise their arm, tilt their head and stare) were judged more dishonest than people who did not show this behaviour (Bond et al., 1992). Also, the more police officers reported to rely on nonverbal cues, the less accurate they were when detecting deceit (Porter, McCabe, Woodworth, & Peace, 2007; Vrij & Mann, 2001). Furthermore, Wachi et al. (2017) showed that interviewers who correctly classified guilty participants, were less likely to report having relied on nonverbal cues. Moreover, deliberately making nonverbal cues inaccessible (e.g. using transcripts or audio fragments) increased deception detection accuracy (Burgoon, Blair, & Strom, 2008; Davis, Markus, & Walters, 2006; Maier & Thurber, 1968; Mann, Vrij, Fisher, & Robinson, 2008). Thus, there seems to be some evidence that stereotypical nonverbal cues do influence credibility judgements.

To investigate whether stereotypical cues to deception, such as gaze aversion and fidgeting, indeed contribute to credibility judgements, we conducted a study in which we directly manipulated the presence of these cues. To test whether the timing of these manipulations matter, we also manipulated whether these cues were present towards the beginning or the end of a statement.

We take into consideration the timing of these manipulations because research consistently shows human decision making is influenced by various cognitive biases (Risinger, Saks, Thompson, & Rosenthal, 2002). One example of these influences is the primacy effect, which refers to the finding that when making a conclusion, we tend to carry more weight to information assimilated early in the decision process, than later (Nickerson, 1998). This effect, in turn, can be explained by the phenomenon of *confirmation bias* (Findley & Scott, 2006; Jones & Sugden, 2001; Risinger et al., 2002). Usually, when making a decision there is too much information for us to take into account, or to give equal consideration. As a solution to this problem, we pay selective attention to information, especially to information that fit with our expectations about the event in question, while ignoring information that does not fit well. The primacy effect, in light of the confirmation bias, implies that because people form an opinion early in the decision process, this information will have the largest influence on how subsequent information will be acquired and interpreted (Risinger et al., 2002).

Various studies have shown the potentially damaging influence of confirmation bias on investigative and judicial processes in the forensic field (for an overview see Kassin, Dror, & Kukucka, 2013), and lie detection specifically (Ben-Shakhar, Bar-Hillel, Bilu, & Shefler, 1998; Bogaard, Meijer, Vrij, Broers, & Merckelbach, 2014; Elaad, Ginton, & Ben-Shakhar, 1994; Levine, Asada, & Park, 2006). For example, Levine et al. (2006) investigated how foreknowledge about whether the interviewee is lying or not would influence observers' perception about how gaze aversive the interviewees were. Interestingly, observers judged liars as more gaze aversive when they had foreknowledge of their veracity than when veracity was unknown. Thus, peoples' perceived honesty influences how observers judge their behaviour.

Furthermore, confirmation bias does not only play a role when judging people's honesty based on their behaviour, but also based on their verbal accounts. Nahari and Ben-Shakhar (2013) asked participants to judge the credibility of a narrator based on two statements. One statement was rich in detail, while the other was poor in detail. Participants were told that richness in detail is a reliable cue to make credibility judgements and were then presented

with both statements. The only difference between the two groups was the order in which they received the statements, but the total number of details presented to them was identical. Participants who first read the rich text judged the narrator as more credible than participants who first read the poor text. Similar results were found when the experimenters manipulated the order of detail richness within a single statement; a text that started rich in details (i.e. primacy effect), received a higher credibility rating than a text presented in the opposite order (i.e. recency effect). Although this study investigated verbal cues to deception —and detailedness more precisely— a similar timing effect might also be observed for stereotypical nonverbal cues.

In sum, the first goal of the current study is to examine whether the presence of stereotypical nonverbal cues influence subsequent credibility judgements. Second, we investigated whether the primacy and recency effects found by Nahari and Ben-Shakhar (2013) generalise to these stereotypical cues.

## 2 | METHODS

### 2.1 | Participants

Seventy-five psychology students (22 males) participated, aged between 18 and 26 years old (M = 21; SD = 1.60). Participants were recruited through a flyer posted on Facebook and/or by directly contacting students, asking them if they would like to participate. They received 1 course credit point as a reward for partaking. The current study was approved by the Ethical Committee of our university.

### 2.2 | Design

The independent variable was the presence of stereotypical behaviour showed in the videos, which was manipulated in a one-way between subject design (control vs. primacy vs. recency condition). Participants were randomly assigned to one of these conditions. The dependent variable was the average credibility judgement for the videos.

### 2.3 | Materials

*Videos*. The participants in this study rated the veracity of audio-visual video clips. Each of these videos consisted of truthful answers to two questions: 'Can you tell me what you did yesterday?' (Q1) and 'Can you tell me what you did the day before yesterday?' (Q2). Six interviewees were asked to answer these questions truthfully twice: The first time, the interviewees were given no further instruction. The second time, the interviewees were instructed to include stereotypical lying behaviours such as 'gaze aversion', 'using hand and arm movements' and 'body movements'. More precisely, interviewees were told that in the manipulated parts they should 'avert their gaze more, use more hand and arm movements (e.g. self-touching and self-fidgeting such as touching their face, playing with their clothes) and to use more body movements (e.g. shift position) than they would normally do when telling a story'. These behaviours were chosen based on the most commonly reported cues used for detecting deceit (The Global Deception Research Team, 2006) but have not been supported by empirical research (DePaulo et al., 2003). Interviewees included these cues in the beginning (first half) when answering Q1, and towards the end (last half) when answering Q2. This served to create the manipulation of primacy and recency effect (see below).

To select the most suitable videos for the current study, all 12 videos were shown to a group of nine people —all undergraduate psychology students— who were asked to evaluate the sound and video quality, and to check whether any video stood out in any way. Based on their quality judgements, we selected the videos that had the best quality scores and did not stand out. This selection resulted in the videos of four interviewees.

**TABLE 1** Overview of conditions

| | Video 1 | | Video 2 | | Video 3 | | Video 4 | |
|---|---|---|---|---|---|---|---|---|
| Control | Q1 | Q2 | *Q1* | *Q2* | Q1 | Q2 | Q1 | Q2 |
| | Control | Control | *Control* | *Control* | Control | Control | Control | Control |
| Primacy | **Q1** | Q2 | *Q1* | *Q2* | **Q1** | Q2 | **Q1** | Q2 |
| | **Stereotype** | Control | *Control* | *Control* | **Stereotype** | Control | **Stereotype** | Control |
| Recency | Q1 | **Q2** | *Q1* | *Q2* | Q1 | **Q2** | Q1 | **Q2** |
| | Control | **Stereotype** | *Control* | *Control* | Control | **Stereotype** | Control | **Stereotype** |

*Note*: The left panels exemplify the three different conditions and each row illustrates which videos and manipulations were shown to the participants in these conditions. Video 2 (bold) was always a control video without any manipulations and was removed from the analyses. The parts depicted in italics are the parts of the video in which the interviewees displayed stereotypical cues.

Each participant rated four videos, one of each interviewee. We constructed three between subject conditions. Participants in the control condition rated the four videos —with each video containing the answers to Q1 and Q2— that contained no manipulation of stereotypical behaviour. Participants in the primacy conditions rated videos that combined the answer to Q1 with stereotypical behaviour with the answer to Q2 without such behaviour. Participants in the recency condition rated the videos that combined the answer to Q1 without stereotypical behaviour, and the answer to Q2 with such behaviour.

To make sure the manipulation in the clips was not too obvious to the participants, the second interviewee presented was always one without stereotypical cues. We deliberate choose the second position as this meant participants would see a maximum two subsequent manipulated videos. This second video was excluded from the subsequent analysis, because it did not include any behavioural manipulations. Which of the interviewees was depicted at what position was completely balanced over participants to prevent any order effects. Importantly, the participants never saw the same person more than once. The videos varied between 95 and 138 seconds and had an average length of 119 seconds (SD = 20.53). See Table 1 for a visual presentation of our design.

## 2.4 | Procedure

After signing the informed consent, participants read a cover story describing that a laptop was stolen from the lab, and that the interviewees whose credibility they had to judge were possible suspects, as they all have access to the lab. Subsequently, they were randomly allocated to one of three conditions; control, primacy or recency condition. Next, participants were shown the four videos on a computer screen. After every video, participants rated the question 'How credible did you find this statement?' on a six-point Likert scale (1 = *not credible*, 6 = *very credible)* with no neutral point. After the last video, participants were asked the open question 'which cues did you use to make your credibility judgements'. Furthermore, they were asked 'How difficult was it to make these credibility judgements?', 'How motivated were you to judge the credibility of these videos?' and 'How much knowledge do you have about lie detection literature?'. All questions were answered on a six-point Likert scale (1 = *not difficult/motivated* to 6 = *very difficult/motivated*). Lastly, the participants were thanked and debriefed.

## 2.5 | Coding of reported cues

Two raters coded all responses to the question 'which cues did you use to make a credibility judgement' and classified them as either nonverbal or verbal. One rater was the lead researcher and was not blind to the study aims. The

other rater was a research assistant who was blind to the research question and manipulation of this study. Within the nonverbal category, responses were further assigned to specific categories such as speech characteristic (e.g. response latency, voice pitch), facial behaviours (e.g. blushing, gaze aversion), and body movements (e.g. hand and arm movements, moving feet), as in previous research (for the complete list see Akehurst, Kohnken, Vrij, & Bull, 1996; Vrij, Akehurst, & Knight, 2006). Based on the answers of our participants, four cues were added to this list: 'insecure', 'sounds rehearsed/prepared', 'thinking hard' and 'fluent speech'. First, we calculated inter-rater reliability of the two raters for all the coded cues. The two raters had a percentage agreement of 94%, which is sufficient to continue our analyses. Importantly, for the subsequent analyses, a cue was only coded as present when both raters agreed upon its presence. This might result in cues missed; however, it leads to a higher reliability of our reported analyses.

## 2.6 | Results

*Manipulation check*. To check whether the videos with the manipulations actually include more obvious stereotypical behaviour, eight independent raters (blind to the research question and manipulation) rated all the videos. We asked them three questions (a) How obvious did the interviewee avert his/her gaze from the camera/interviewer, (b) How obvious did the interviewee move his/her hands/arm/fingers (e.g. hand and arm movements, touching their face and hair), and (c) How obvious did the interviewee move his/her body (e.g. shifting, turning away from or to the interviewer). All questions were answered on a 10-point scale ranging from 1 (i.e. very subtle) to 10 (i.e. very obvious). As we were interested in the presence of the nonverbal cues combined, we calculated the average scores for all questions combined per condition. The average score for the videos without any manipulation was 4.50 (SD = 1.02), for the primacy videos was 5.51 (SD = .83) and for the recency videos was 5.10 (SD = .95), showing our manipulation was successful.

*Statistical analyses*. To examine our hypotheses, we conducted a repeated measures ANOVA, and a repeated measures Bayes ANOVA with JASP. We report the effect sizes and accompanying Bayesian Factors (BF), as BFs exhibit better performance than the *p*-value for testing the null hypothesis (García & Puga, 2018). Evidence for the interaction model is calculated as [interaction model]/[main factors] (see Wagenmakers et al., 2016). Bayes Factors allow three types of conclusions: evidence for H1 (BF > 1); evidence for H0 (BF close to 0); and the evidence is insensitive (BF close to 1; Dienes, 2014). The approximate classification scheme is the following: BF > 100 extreme evidence for H1, 30–100 very strong evidence, 10–30 strong evidence; 3–10 moderate evidence and 1–3 anecdotal

**TABLE 2** Means and standard deviations of self-reported motivation, difficulty and knowledge of lie detection literature

| | Motivation | | Difficulty | | Knowledge | |
|---|---|---|---|---|---|---|
| Condition | M | SD | M | SD | M | SD |
| Control | 4.64 | 0.91 | 3.52 | 1.48 | 2.68 | 1.28 |
| Primacy | 4.69 | 1.19 | 3.46 | 1.30 | 2.54 | 1.30 |
| Recency | 4.54 | 0.83 | 4.00 | 1.18 | 2.42 | 1.28 |

**TABLE 3** Means and standard deviations of credibility scores separated per condition

| | Video 1 | | Video 3 | | Video 4 | | Average | |
|---|---|---|---|---|---|---|---|---|
| Condition | M | SD | M | SD | M | SD | M | SD |
| Control | 4.36 | 1.11 | 4.28 | 1.49 | 4.36 | 1.08 | 4.33 | 0.92 |
| Primacy | 3.50 | 1.10 | 3.19 | 1.33 | 4.04 | 1.51 | 3.57 | 0.67 |
| Recency | 3.92 | 1.21 | 3.29 | 1.52 | 3.25 | 1.22 | 3.84 | 0.82 |

evidence for H1 (for further interpretation of BF see Jarosz & Wiley, 2014; Lee & Wagenmakers, 2013). For ease of interpretation, $BF_{10}$ is used to indicate evidence in favour of $H_1$, whereas $BF_{01}$ is used to indicate evidence in favour of $H_0$.

**TABLE 4** Frequencies and percentages for the verbal and nonverbal cues that were mentioned, separated per condition

| | Control | | Primacy | | Recency | |
|---|---|---|---|---|---|---|
| **Cues** | **Frequency** | **Percentage** | **Frequency** | **Percentage** | **Frequency** | **Percentage** |
| **Non-verbal cues** | | | | | | |
| Non-verbal behaviour | 12 | 17.39 | 13 | 16.25 | 13 | 26.00 |
| Eye contact | 12 | 17.39 | 14 | 17.50 | 4 | 8.00 |
| Insecure | 4 | 5.80 | 1 | 1.25 | 1 | 2.00 |
| Facial expressions | 3 | 4.35 | 2 | 2.50 | 1 | 2.00 |
| Thinking | 3 | 4.35 | 1 | 1.25 | 1 | 2.00 |
| Self-manipulations | 3 | 4.35 | 0 | 0.00 | 0 | 0.00 |
| Faltering speech | 2 | 2.90 | 5 | 6.25 | 0 | 0.00 |
| Hand/arm/finger movement | 1 | 1.45 | 3 | 3.75 | 2 | 4.00 |
| Sounds rehearsed/ prepared | 1 | 1.45 | 0 | 0.00 | 2 | 4.00 |
| Nervous behaviour | 1 | 1.45 | 1 | 1.25 | 1 | 2.00 |
| Pauses | 1 | 1.45 | 2 | 2.50 | 0 | 0.00 |
| Gestures | 1 | 1.45 | 2 | 2.50 | 0 | 0.00 |
| Smiling | 1 | 1.45 | 0 | 0.00 | 0 | 0.00 |
| Postural shifts | 1 | 1.45 | 0 | 0.00 | 0 | 0.00 |
| Clearing throat | 0 | 0.00 | 2 | 2.50 | 0 | 0.00 |
| **Speech cues** | | | | | | |
| Speech | 5 | 7.25 | 6 | 7.50 | 1 | 2.00 |
| Monotonous speech | 1 | 1.45 | 0 | 0.00 | 0 | 0.00 |
| Fluent speech | 0 | 0.00 | 1 | 1.25 | 3 | 6.00 |
| Hectic speech | 0 | 0.00 | 3 | 3.75 | 2 | 4.00 |
| **Verbal cues** | | | | | | |
| Amount of details | 10 | 14.49 | 9 | 11.25 | 8 | 16.00 |
| Verbal cues | 4 | 5.80 | 6 | 7.50 | 3 | 6.00 |
| Inconsistency | 1 | 1.45 | 2 | 2.50 | 3 | 6.00 |
| Temporal details | 1 | 1.45 | 1 | 1.25 | 0 | 0.00 |
| Emotions | 1 | 1.45 | 0 | 0.00 | 0 | 0.00 |
| Plausibility | 0 | 0.00 | 0 | 0.00 | 2 | 4.00 |
| Coherence | 0 | 0.00 | 2 | 2.50 | 1 | 2.00 |
| Chronological production | 0 | 0.00 | 2 | 2.50 | 1 | 2.00 |
| Structure | 0 | 0.00 | 0 | 0.00 | 1 | 2.00 |
| Repetition | 0 | 0.00 | 2 | 2.50 | 0 | 0.00 |

*Pre-analyses check*. We first checked whether there were any gender effects. One person failed to provide this information and could not be taken into consideration. Results showed no gender main effect [$F(1,74) < 0.00$, $p = 0.99$, $n_p^2 < .001$] and no gender x condition interaction effect [$F(2,74) = 0.43$, $p = 0.65$, $n_p^2 = .01$].

To examine whether the groups differed on motivation, perceived task difficulty, and background knowledge of deception, we conducted three one-way (Bayes) ANOVAs. The analyses showed no difference in motivation [$F(2,74) = .15$, $p = 0.86$, $BF_{10} = .29$], perceived task difficulty [$F(2,74) = 1.22$, $p = 0.30$, $BF_{10} = .12$], or background knowledge of lie detection [$F(2,74) = 0.26$, $p = 0.78$, $BF_{10} = .14$] between conditions (see Table 2).

*Credibility*. To assess how the presence of nonverbal stereotypical cues influence credibility judgements we calculated an overall credibility score for the videos in all three conditions, with the exclusion of the second video as this was a video without behavioural manipulations. The second video was always a control video without manipulations to make early and late onset of stereotypical cues less obvious to the participants. So, we do not expect any differences in credibility scores for video 2. To check this, we ran a one-way ANOVA on the scores for video 2. As expected, results showed no significant differences between groups [$F(2,74) = 1.09$, $p = 0.34$, $n_p^2 = .03$]. Given the study aims to investigate the influence of stereotypical cues, we only take the three manipulated videos into consideration for the following analyses.

We conducted a repeated measures (Bayes) ANOVA and included the credibility scores of the three videos as the within subject factor, and condition as the between subject factor. As hypothesized, results showed a significant difference in credibility scores between conditions [$F(2,72) = 8.21$, $p = 0.001$], strongly supported by the Bayes Factor ($BF_{10} = 20.26$). No differences between videos [$F(2,71) = 1.89$, $p = 0.16$, $BF_{10} = 4.80$] or a video x condition interaction effect emerged [$F(4,142) = 1.61$, $p = 0.18$, $BF_{01} = 2.27$]. Subsequent post-hoc analyses showed that participants in both the primacy and the recency condition rated the credibility of the videos significantly lower than the control condition (Bonferroni corrected). The mean difference (*MD*) between the control condition and the primacy condition was $MD = -.76$ ($p = 0.004$; $BF_{10,U} = 61.89$). The mean difference between the control condition and the recency condition was $MD = -.85$ ($p = 0.001$; $BF_{10,U} = 213.44$). No significant difference was found between the primacy and recency condition ($MD = 0.10$), which was supported by the Bayes Factor ($BF_{01,U} = 5.26$). See Table 3 for an overview of means and standard deviations per condition.

*Cues used to make a judgement*. To our question 'Which cues did you use to make a credibility judgement?', participants gave a total of 199 different responses: 141 nonverbal and 58 verbal cues. An overview of all mentioned cues is provided in Table 4. As expected, the manipulated stereotypical nonverbal cues were also mentioned most often. The most common nonverbal cues in the control condition were (a) eye contact, (b) behaviour (not further specified) and (c) speech. For the primacy group these were (a) eye contact, (b) behaviour (not further specified) and (c) speech. For the recency group these were (a) behaviour (not further specified), (b) eye contact and (c) fluent speech. Participants only mentioned nine different verbal cues, and the three most common verbal cues were similar in all groups, namely (a) amount of details, (b) verbal cues (not further specified) and (c) inconsistencies/contradictions.

## 2.7 | Discussion

Overall, many people report to rely on nonverbal cues to assess credibility (The Global Deception Research Team, 2006). With the current study we aimed to bypass many problems associated with self-reported insights into behaviour (Nisbett & Wilson, 1977), therefore we directly manipulated the presence and timing of stereotypical deception cues, and investigated to what extent their presence influenced credibility judgements. Our results confirmed our primary hypothesis that the greater presence of stereotypical behavioural cues significantly reduced the perceived credibility of the interviewees. However, in contrast to our expectations, the timing of these cues did not influence our participants' credibility judgements.

Our finding that stereotypical cues contribute to credibility judgements contrasts the findings of Hartwig and Bond (2011). Besides being reflected in the difference in credibility judgements, the validity of our finding is further reinforced by the cues people reported to have used for their decision. This showed that participants reported to indeed rely on the nonverbal cues we manipulated. Also, participants listed considerably more nonverbal than verbal cues as diagnostic, findings that are in correspondence with previously published deception studies (e.g. Akehurst et al., 1996; Bogaard et al., 2016; Masip & Herrero, 2015; Strömwall & Granhag, 2003; Strömwall et al., 2004; Taylor & Hick, 2007; Vrij, Akehurst, et al., 2006; Vrij & Semin, 1996).

The importance of the finding that the greater presence of stereotypical cues influence credibility judgements can be illustrated when looking at the process of behavioural confirmation bias in the investigative process (Kassin, Goldstein, & Savitsky, 2003). These (wrongful) credibility judgements can lead to severe behavioural consequences in social interactions (Darley & Fazio, 1980; Kassin et al., 2013). For example, turning to a harsher interrogation style, or asking more guilt presumptive questions when an interviewee is expected to be guilty (Hill, Memon, & McGeorge, 2008; Kassin et al., 2003; Narchet, Meissner, & Russano, 2011). Based on these findings we can assume that showing stereotypical 'deceptive' behaviour during an interrogation —early or late— negatively influences the investigative process. Indeed, Novotny et al. (2018) reported that behavioural cues are often relied upon in the early stage of the lie detection process and sets in motion the search for other evidence. This, in turn, aligns with Kassin's (2015) assumption that the starting point of false confessions and wrongful convictions is often mistaken lie detection by investigators who focus too much on stereotypical cues. Finally, research on cultural differences and nonverbal behaviour during interviews has shown that some ethnic groups (e.g. Surinamese) naturally demonstrate more cues commonly associated with deception (i.e. more gaze aversion; more trunk, hand and arm movements) when truth telling (Vrij & Winkel, 1991). Our results imply that such behaviour can be misinterpreted by police officers as signs of deceit, and also set into motion the process of behavioural confirmation bias.

We did not find a nonverbal primacy effect, which contrasts the findings by Nahari and Ben-Shakhar (2013). These authors showed a clear primacy effect for the verbal cue richness in details. A possible explanation for this discrepancy may lie in the instructions participants received just prior to reading the statements in Nahari and Ben-Shakhar (2013). In that study, participants were explicitly informed that richness is a reliable cue for truthfulness. This may have led participants to search for these cues especially in the beginning of the statement. We did not explicitly instruct our participants about the usefulness of cues to deception before viewing the videos, as such instructions would invalidate our primary research question. We assumed they would automatically infer that 'gaze aversion', 'using hand and arm movements' and 'body movements' indicate deception. A belief our participants indeed reported to endorse, showing these cues are a proper manipulation of stereotypical cues to deception.

Hartwig and Bond (2011) concluded from their meta-analytical findings that people rarely rely on invalid cues when making honesty judgements. However, our findings show there are exceptions to this conclusion. A possible explanation for these contrasting findings might lie in how many useful cues are available to the observer. Thus, the strength in which people rely on stereotypical cues may depend on the number of available objective cues. Perhaps, if there are only few cues available (e.g. because the interviewee is reluctant to give information), interviewers focus more on visual cues as these are always readily available. Our data provides partial support for this availability hypothesis. Participants did report having used most non-verbal cues in the primacy condition (57 cues in total) but less non-verbal cues in the recency condition (37 cues in total) than in the control condition (52 cues in total). Why we did not find support for this hypothesis in the recency condition, we can only speculate. Although the clips were relatively short, participants might have had difficulties remembering all the cues portrayed, especially towards the end. Later information is usually less deeply processed than early information (Krosnick & Alwin, 1987) and may therefore be reported less.

Furthermore, the cue availability assumption might also explain why high-stake lies are sometimes more accurately detected than low-stake lies (Mann, Vrij, & Bull, 2002, 2004; Vrij & Mann, 2001; Wachi et al., 2017; Wright Whelan, Wagstaff, & Wheatcroft, 2014, 2015; but see Hartwig & Bond, 2014). Increasing the stakes might intensify the production of cues to deception, causing liars to exhibit more noticeable/useful cues as a result of increased

feelings of fear, guilt and cognitive load (Mann et al., 2004; Vrij, 2008; Vrij, Mann, Robbins, & Robinson, 2006; Wright Whelan et al., 2015). For example, Mann et al. (2002) showed that high stake liars blinked less frequently and made longer pauses, but did not differ in stereotypical cues usually associated with deception. However, in the current study, we used alibi statements and the emotional valence and/or content might be limited. Consequently, also the number of available objective cues our participants could use. Hence, they relied on stereotypical—yet invalid—visual cues. Based on these findings, it seems that people are flexible in which cues they rely on when detecting deception, depending on cue availability. However, the current dataset does not allow us to test this assumption.

Two limitations deserve attention. First, we asked participants to answer the questions naturally first, and then to repeat their answer and include stereotypical cues. One may argue that this causes the presence and absence of the manipulated stereotypical cues to be confounded by rehearsal. The literature indeed hints that judges who have the impression that a statement is rehearsed rate these statements lower in credibility (Landström, Granhag, & Hartwig, 2005). This means the effects of rehearsal can explain our pattern of results. On the other hand, meta-analytic research has shown that planned and/or rehearsed lies result in fewer observable cues to deception than spontaneous lies (DePaulo et al., 2003). This would mean our findings provide an underestimation of the true effect. This latter explanation is supported by the self-report data: only three out of the 75 participants (one in the control, two in the recency condition) reported to have used the 'sounds rehearsed/prepared' cue for their credibility ratings. Second, we investigated primacy and recency effects clips that are shorter than the typical police interview. As a result, one can question to what extent these effects translate to longer interviews.

In sum, our results document that the presence of behavioural cues can decrease the observed credibility of a narrator, and regardless of whether these stereotypical cues are present in the beginning, or the end of the statement. Despite research showing that there is a discrepancy between self-reported cues and cues used when making honesty judgements (Hartwig & Bond, 2011), the current findings shows this discrepancy does not hold for some stereotypical cues.

## ORCID

*Glynis Bogaard* https://orcid.org/0000-0001-6795-9433

## REFERENCES

Akehurst, L., Kohnken, G., Vrij, A., & Bull, R. (1996). Lay persons' and police officers' beliefs regarding deceptive behaviour. *Applied Cognitive Psychology*, *10*(6), 461–471. https://doi.org/10.1002/(Sici)1099-0720(199612)10:6<461::Aid-Acp413>3.0.Co;2-2

Ben-Shakhar, G., Bar-Hillel, M., Bilu, Y., & Shefler, G. (1998). Seek and you shall find: A confirmation bias in clinical judgment. *Journal of Behavioral Decision Making*, *11*, 235–249.

Bogaard, G., Meijer, E. H., Vrij, A., Broers, N. J., & Merckelbach, H. (2014). Contextual bias in verbal credibility assessment: Criteria-based content analysis, reality monitoring and scientific content analysis. *Applied Cognitive Psychology*, *28*(1), 79–90. https://doi.org/10.1002/acp.2959

Bogaard, G., Meijer, E. H., Vrij, A., & Merckelbach, H. (2016). Strong, but wrong: Lay people's and police officers' beliefs about verbal and nonverbal cues to deception. *PLoS One*, *11*(6), e0156615. https://doi.org/10.1371/journal.pone.0156615

Bond, C. F., Omar, A., Pitre, U., Lashley, B. R., Skaggs, L. M., & Kirk, C. (1992). Fishy-looking liars: Deception judgment from expectancy violation. *Journal of Personality and Social Psychology*, *63*(6), 969–977.

Burgoon, J. K., Blair, J. P., & Strom, R. E. (2008). Cognitive biases and nonverbal cue availability in detecting deception. *Human Communication Research*, *34*(4), 572–599. https://doi.org/10.1111/j.1468-2958.2008.00333.x

Darley, J. M., & Fazio, R. H. (1980). Expectancy confirmation processes arising in the social interaction sequence. *American Psychologist*, *35*(10), 867–881. https://doi.org/10.1037/0003-066x.35.10.867

Davis, M., Markus, K. A., & Walters, S. B. (2006). Judging the credibility of criminal suspect statements: Does mode of presentation matter? *Journal of Nonverbal Behavior*, *30*, 181–198.

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.

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, *5*, 781. https://doi.org/10.3389/fpsyg.2014.00781

Elaad, E., Ginton, A., & Ben-Shakhar, G. (1994). The effects of prior expectations and outcome knowledge on polygraph examiners' decisions. *Journal of Behavioral Decision Making*, *7*(4), 279–292. https://doi.org/10.1002/bdm.3960070405

Findley, K. A., & Scott, M. S. (2006). The multiple dimensions of tunnel vision in criminal cases. *Wisconsin Law Review*, *2*(2), 291–397.

García, A. M. R.-R., & Puga, J. L. (2018). Deciding on null hypotheses using P-values or Bayesian alternatives: A simulation study. *Psicothema*, *30*(1), 110–115.

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

Hartwig, M., & Bond, C. F. (2014). Lie Detection from Multiple Cues: A Meta-analysis. *Applied Cognitive Psychology*, *28*(5), 661–676. https://doi.org/10.1002/acp.3052

Hartwig, M., & Granhag, P. A. (2014). Exploring the nature and origin of beliefs about deception: implicit and explicit knowledge among lay people and presumed experts. In P. A. Granhag, A. Vrij, & B. Verschuere (Eds.), *Detecting deception: Current challenges and cognitive approaches* (pp. 125–154). Chichester: John Wiley & Sons.

Hill, C., Memon, A., & McGeorge, P. (2008). The role of confirmation bias in suspect interviews: A systematic evaluation. *Legal and Criminological Psychology*, *13*, 357–371. https://doi.org/10.1348/135532507x238682

Jarosz, A. F., & Wiley, J. (2014). What are the odds? A practical guide to computing and reporting Bayes factors. *The Journal of Problem Solving*, *7*, 2–9.

Jones, M., & Sugden, R. (2001). Positive confirmation bias in the acquisition of information. *Theory and Decision*, *50*(1), 59–99. https://doi.org/10.1023/A:1005296023424

Kassin, S. M. (2015). The social psychology of false confessions. *Social Issues and Policy Review*, *9*(1), 25–51. https://doi.org/10.1111/sipr.12009

Kassin, S. M., Dror, I. E., & Kukucka, J. (2013). The forensic confirmation bias: Problems, perspectives, and proposed solutions. *Journal of Applied Research in Memory and Cognition*, *2*(1), 42–52. https://doi.org/10.1016/j.jarmac.2013.01.001

Kassin, S. M., Goldstein, C. C., & Savitsky, K. (2003). Behavioral confirmation in the interrogation room: On the dangers of presuming guilt. *Law and Human Behavior*, *27*(2), 187–203.

Krosnick, J. A., & Alwin, D. F. (1987). An evaluation of a cognitive theory of response-order effects in survey measurement. *Public Opinion Quarterly*, *51*(2), 201–219.

Landström, S., Granhag, P. A., & Hartwig, M. (2005). Witnesses appearing live versus on video: Effects on observers' perception, veracity assessments and memory. *Applied Cognitive Psychology*, *19*(7), 913–933.

Lee, M. D., & Wagenmakers, E. J. (2013). *Bayesian cognitive modeling: A practical course*. Cambridge, UK: Cambridge University Press.

Levine, T. R., Asada, K. J. K., & Park, H. S. (2006). The lying chicken and the gaze avoidant egg: Eye contact, deception, and causal order. *Southern Communication Journal*, *71*(4), 401–411. https://doi.org/10.1080/10417940601000576

Maier, N. R. F., & Thurber, J. A. (1968). Accuracy of judgments of deception when an interview is watched, heard, and read. *Personnel Psychology*, *21*, 23–30.

Mann, S., Vrij, A., & Bull, R. (2002). Suspects, lies and videotape: An analysis of authentic high-stakes liars. *Law and Human Behavior*, *26*(3), 365–376. https://doi.org/10.1023/A:1015332606792

Mann, S., Vrij, A., & Bull, R. (2004). Detecting true lies: Police officers' ability to detect suspects' lies. *The Journal of Applied Psychology*, *89*(1), 137–149. https://doi.org/10.1037/0021-9010.89.1.137

Mann, S. A., Vrij, A., Fisher, R. P., & Robinson, M. (2008). See no lies, hear no lies: Differences in discrimination accuracy and response bias when watching or listening to police suspect interviews. *Applied Cognitive Psychology*, *22*(8), 1062–1071. https://doi.org/10.1002/acp.1406

Masip, J., & Herrero, C. (2015). Police detection of deception: Beliefs about behavioral cues to deception are strong even though contextual evidence is more useful. *Journal of Communication*, *65*(1), 125–145. https://doi.org/10.1111/jcom.12135

Nahari, G., & Ben-Shakhar, G. (2013). Primacy effect in credibility judgements: The vulnerability of verbal cues to biased interpretations. *Applied Cognitive Psychology*, *27*(2), 247–255. https://doi.org/10.1002/acp.2901

Narchet, F. M., Meissner, C. A., & Russano, M. B. (2011). Modeling the influence of investigator bias on the elicitation of true and false confessions. *Law and Human Behavior*, *35*(6), 452–465. https://doi.org/10.1007/s10979-010-9257-x

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*, 175–220.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, *84*(3), 231–259.

Novotny, E., Carr, Z., Frank, M. G., Dietrich, S., Shaddock, T., Cardwell, M., & Decker, A. (2018). How people really suspect and discover lies. *Journal of Nonverbal Behavior*, *42*(1), 41–52. https://doi.org/10.1007/s10919-017-0263-2

Porter, S., McCabe, S., Woodworth, M., & Peace, K. A. (2007). Genius is 1% inspiration and 99% perspiration… or is it? An investigation of the impact of motivation and feedback on deception detection. *Legal and Criminological Psychology*, *12*, 297–309.

Risinger, D. M., Saks, M. J., Thompson, W. C., & Rosenthal, R. (2002). The Daubert/Kumho implications of observer effects in forensic science: Hidden problems of expectation and suggestion. *California Law Review*, *90*(1), 1–56. https://doi.org/10.2307/3481305

Strömwall, L. A., & Granhag, P. A. (2003). How to detect deception? Arresting the beliefs of police officers, prosecutors and judges. *Psychology, Crime & Law*, *9*, 19–36.

Strömwall, L. A., Granhag, P. A., & Hartwig, M. (2004). Practicioners' beliefs about deception. In L. A. Strömwall & P. A. Granhag (Eds.), *The detection of deception in forensic contexts*. Cambridge, UK: Cambridge University Press.

Taylor, R., & Hick, R. F. (2007). Believed cues to deception: Judgments in self-generated trivial and serious situations. *Legal and Criminological Psychology*, *12*, 321–331. https://doi.org/10.1348/135532506x116101

The Global Deception Research Team. (2006). A world of lies. *Journal of Cross-Cultural Psychology*, *37*, 60–74.

Vrij, A. (2008). *Detecting lies and deceit: Pitfalls and opportunities*. Chichester, England: Wiley.

Vrij, A., Akehurst, L., & Knight, S. (2006). Police officers', social workers', teachers' and the general public's beliefs about deception in children, adolescents and adults. *Legal and Criminological Psychology*, *11*, 297–312. https://doi.org/10.1348/135532505x60816

Vrij, A., & Mann, S. (2001). Who killed my relative? Police officers' ability to detect real-life high stake lies. *Psychology, Crime & Law*, *7*, 119–132.

Vrij, A., Mann, S., Robbins, E., & Robinson, M. (2006). Police officers ability to detect deception in high stakes situations and in repeated lie detection tests. *Applied Cognitive Psychology*, *20*(6), 741–755. https://doi.org/10.1002/acp.1200

Vrij, A., & Semin, G. (1996). Lie experts' beliefs about nonverbal indicators of deception. *Journal of Nonverbal Behavior*, *20*, 65–80.

Vrij, A., & Winkel, F. W. (1991). Cultural patterns in Dutch and Surinam nonverbal behavior: An analysis of simulated police/citizen encounters. *Journal of Nonverbal Behavior*, *15*(3), 169–184. https://doi.org/10.1007/Bf01672219

Wachi, T., Kuraishi, H., Watanabe, K., Otsuka, Y., Yokota, K., & Lamb, M. E. (2017). Police officers' ability to detect lies within a deception paradigm. *Psychology, Public Policy, and Law*, *23*(3), 301–311. https://doi.org/10.1037/law0000124

Wagenmakers, E. J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., … Morey, R. D. (2016). Bayesian Inference for Psychology. Part II: Example Applications with JASP. *Psychonomic Bulletin & Review*, *25*(1), 58–76. https://doi.org/10.3758/s13423-017-1323-7

Wright Whelan, C., Wagstaff, G., & Wheatcroft, J. M. (2015). High stakes lies: Police and non-police accuracy in detecting deception. *Psychology, Crime & Law*, *21*, 127–138.

Wright Whelan, C., Wagstaff, G. F., & Wheatcroft, J. M. (2014). High-stakes lies: Verbal and nonverbal cues to deception in public appeals for help with missing or murdered relatives. *Psychiatry, Psychology and Law*, *21*(4), 523–537.

**How to cite this article:** Bogaard G, Meijer EH. Stereotypical behavioural cues — but not their order — influence credibility judgements. *J Investig Psychol Offender Profil*. 2020;17:131–141. https://doi.org/10.1002/jip.1543