Does (Biasing) Nonverbal Information Deteriorate the Accuracy of the Take-the-Best Heuristic for Deception Detection?

Lob, A.B.; Chakir, N.; van Munster van Heuven, L.; Verschuere, B.

DOI
10.1002/acp.70006

Publication date
2024

Document Version
Final published version

Published in
Applied Cognitive Psychology

License
CC BY-NC-ND

Applied Cognitive Psychology

WILEY

RESEARCH ARTICLE OPEN ACCESS

Does (Biasing) Nonverbal Information Deteriorate the Accuracy of the Take-the-Best Heuristic for Deception Detection?

Aaron Benjamin Lob[1] | Nisrin Chakir[2] | Laurine van Munster van Heuven[2] | Bruno Verschuere[2]

[1]Cognitive and Behavioral Decision Research, Department of Psychology, University of Zurich, Zurich, Switzerland | [2]Department of Clinical Psychology, University of Amsterdam, Amsterdam, The Netherlands

Correspondence: Bruno Verschuere (b.j.verschuere@uva.nl)

Received: 2 August 2024 | **Revised:** 15 October 2024 | **Accepted:** 24 October 2024

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

Keywords: boosting | decision making | Detailedness | heuristics | honesty | lie detection

ABSTRACT

People are poor lie detectors, partly because they hold false beliefs about nonverbal cues to deception. Here, we investigated if guiding people to rely only on a message's detailedness ("take-the-best") boosts their lie detection and to what extent such heuristic judgments are immune to nonverbal information. In three studies (*N*s = 109, 88 and 144), participants made detailedness-based veracity judgements, of text versus video statements (Study 1), or of statements without or with biasing nonverbal behavior (truth tellers diverting, liars maintaining gaze; Studies 2 and 3). Compared to unguided judgements, participants using the heuristic method achieved higher deception detection accuracy throughout. Mere access to nonverbal behavior did not deteriorate performance (Study 1), but the heuristic was not fully immune to biasing nonverbal behavior (Studies 2, 3). Our findings challenge the lay notion that access to nonverbal behavior benefits deception detection and suggest that only focusing on diagnostic cues improves lie detection.

1 | Introduction

A spectacular example of how people can be misguided by charismatic individuals' nonverbal behavior is the case of Elizabeth Holmes, founder of *Theranos*. Once an admired pioneer in the field of blood test diagnostics, Holmes was convicted of wire fraud after reports that her method did in fact not work. Yet, for a long time, Holmes managed to convince patients, journalists, and investors of the opposite. Holmes' charisma is well-documented and includes her habit of never breaking eye contact to convince the public of her method and show her credibility (McShane 2019; Popken 2021; Way 2021). Cases such as the one of Elizabeth Holmes and TV shows such as "Lie to me" show that many people hold beliefs about what nonverbal behavior can and should be used if one wants to detect deceit. Yet, nonverbal behavior in general—and a person's eye gaze in particular—is repeatedly found to be not useful when trying to evaluate someone's truthfulness (e.g., DePaulo et al. 2003). Thus, the question remains, whether people are, although potentially well-intentioned, simply focusing on the wrong cues when trying to tell lie from truth, thus rendering their judgments worse than they could be?

There is substantial evidence for the assumption that people are poor lie detectors. Meta-analyses find that laypeople perform only slightly above chance level (i.e., 54%; Bond and

The preregistration for Study 1 is available here: https://aspredicted.org/blind.php?x=QP8_XFY, the preregistration for Study 2 is available here: https://aspredicted.org/PZ5_VS8, and the preregistration for Study 3 is available here: https://aspredicted.org/tn89r.pdf. Material, data, analysis scripts, and supplementary material that support the findings of this article are publicly available on OSF: https://osf.io/7wtyz/.

DePaulo 2006). Even police officers and judges do not perform better (Aamodt and Custer 2006; Akehurst et al. 1996; Leach et al. 2004). Why is that? A person can use a lot of accessible cues for lie detection. However, DePaulo et al. (2003) show that only a few out of 158 potential cues are (typically weakly) related to deception (see also, Bogaard et al. 2016). Thus, people face the problem of choosing which of the many cues to consider and how to integrate their implications into one judgment. Content-related lie-detection tools used by trained coders (e.g., CBSA, Steller and Köhnken 1989; Nahari, Vrij, and Fisher 2014) make use of certain subsets of cues and allow for above-chance accuracy in lie detection. Yet, these methods are resource-intense and not suitable for situations in which practitioners need to make quick decisions. With the current research, we want to contribute to a refined understanding of how less resource-intense methods, that is, heuristic deception detection methods (Verschuere et al. 2023), work in improving deception detection. By helping people to focus on the right cues when trying to tell lie from truth, these methods might prove to be useful tools for laypeople and experts alike.

2 | Heuristics: A Shortcut to Deception Detection

The idea behind the heuristic approach to improving people's deception judgments is to provide simple decision rules (Gigerenzer and Todd 1999; Tversky and Kahneman 1974). A heuristic is a decision rule that simplifies a complex task (Tversky and Kahneman 1974) by reducing the amount of information used or needed to make a decision (Gigerenzer and Todd 1999). Thus, they allow people to make judgments with little available time and resources. An example of such a resource-efficient decision strategy is "Take-the-best" (Gigerenzer and Goldstein 1996). Take-the-best works according to the policy "take the best, ignore the rest", ranking cues by their diagnostic value and only using the cue with the highest validity (i.e., correlation with the criterion) for making decisions. Despite thorough mapping of how they can err (Kahneman et al. 1982; Tversky and Kahneman 1974), it has been repeatedly shown that heuristics can be surprisingly accurate, performing equally, or better than resource-intense decision algorithms in domains such as prediction of criminal recidivism (Dressel and Farid 2018), life outcomes (Salganik et al. 2020), and medical decisions (Gigerenzer and Todd 1999). Could focusing on an individual, highly diagnostic cue (i.e., following a take-the-best heuristic) also boost laypeople's performance in deception detection?

Theoretically, using one diagnostic cue for deception detection should help in better identifying deceit. The reason for this is simple: Since detailedness is the best available cue to deception (DePaulo et al. 2003; Luke 2019; Verschuere, Bogaard, and Meijer 2021), focusing exclusively on a statement's degree of detailedness should outperform an unguided judgment. In the latter case, people might focus on a number of cues with varying degree of diagnostic value, rendering judgments based on this bundle of cues less well attuned to the actual truthfulness of a statement. Of course, this relationship hinges upon the assumption that truthful and deceitful statements are differently associated with their degree of detailedness.

In recent years, a number of research lines combined show that detailedness is the most diagnostic cue and perhaps the only cue for which there exists robust empirical evidence. First, based on extensive simulation studies, Luke (2019) showed that publication bias, small numbers of estimates, and low power inflate the effect size estimates of cues to deception, with detailedness being the sole exception. Second, as far as we are aware, detailedness is the only cue that is supported by independent preregistered replications (Verschuere et al. 2021). Third, meta-analytic research confirmed the validity of detailedness as a cue to deception (DePaulo et al. 2003) and has indicated its effect size can be substantial under certain conditions (i.e., the alibi scenario: $d = 0.94$ [95% CI: 0.64, 1.25]; Verschuere, Bogaard, and Meijer 2021). Fourth, a survey of deception experts showed that the highest level of consensus (ca. 70%) on cues predictive of deception was reached for detailedness (Luke 2019). But not only empirical evidence suggests that detailedness is associated with a statement's truthfulness. Reality monitoring theory provides a strong theoretical argument for this, holding that characteristics of a memory may inform on its origin (Johnson and Raye 1981). This means that truth tellers, having experienced the event, can provide an account that is richer in detail than liars (Masip et al. 2005). Moreover, liars risk being exposed when inventing details and may therefore refrain from providing them (Hartwig and Bond 2014; Nahari 2018).

Indeed, Verschuere et al. (2023) found that heuristic judgments relying only on detailedness allowed lay people to distinguish between truths and lies, outperforming not only unguided judgments where participants were free to use any possible cue but also judgments based on several cues. Of note, in these studies, participants evaluated text statements and had no access to nonverbal behavior. In real-life situations, however, people will usually see and hear the person whose statement they evaluate, confronting them with a number of presumably nondiagnostic nonverbal cues.

3 | The Influence of (Biasing) Nonverbal Behavior

The question is thus whether and how nonverbal behavior impacts heuristic-based veracity judgements in such situations. There is substantial evidence suggesting that it does—in a negative way. People have a strongly stereotypical image of how liars behave (Bond and DePaulo 2006). For example, nonverbal behavior such as whether someone seems credible is one of the most predictive factors for whether people believe them or not (Bond and DePaulo 2008). The most mentioned nonverbal cue, however, is that liars avert their eye gaze (Global Deception Research Team 2006; Sporer and Schwandt 2007; Taylor and Hick 2007). Gaze aversion, however, is not objectively associated with deception ($d = 0.03$; DePaulo et al. 2003). Thus, while judges have deemed access to nonverbal behavior critical for credibility assessment (Leach et al. 2016), it may actually deter from focusing on diagnostic cues and instead focus on nondiagnostic cues such as a person's gaze aversion. Consequently, Vrij and Hartwig (2021) argued that making nonverbal information unavailable (e.g., by covering part of the face) may increase deception detection since nondiagnostic nonverbal information cannot interfere. Indeed, Burgoon, Blair, and Strom (2008) found that the more visual information is available, the worse and more biased deception judgments were. In an experimental study, participants' deception detection accuracy reduced when visual information of senders giving a statement (i.e., video tapes) was available, as opposed to when only audio recordings were available. Similarly, a meta-analysis by Bond and DePaulo (2006) shows that people perform worse in lie detection when judging the veracity of visual content as compared to audible content. Finally, laypeople's deception detection accuracy increased when evaluating statements given by females wearing veils compared to those not wearing veils (Leach et al. 2016; for a commentary see Denault et al. 2017) although, in contrast with the idea that shielding of nonverbal behavior would benefit lie detection, they found that accuracy was higher for women wearing a hijab than those wearing a niqab (while that covers more of the face and body). Irrespective of these, in all reported studies, people made intuitive judgements.

Here, we compared such unguided judgments with judgments made by people guided to use only the level of detail (i.e., make heuristic judgments). Thus, the "detailedness heuristic" introduced here relies on the same principles as other take-the-best heuristics discussed above, that is, basing one's judgments only on the most diagnostic cue. Openly instructing people to use diagnostic information for their judgments can be considered a "boosting intervention" (Hertwig and Grüne-Yanoff 2017; Lorenz-Spreen et al. 2021). Following this, by guiding people to focus only on detailedness, we aimed to help lay people to avoid relying on nondiagnostic nonverbal behavior and thus improve their deception detection accuracy. Study 1 examines the influence of mere access to nonverbal behavior (video vs. text), and Studies 2 and 3 examine the influence of strongly biasing nonverbal behavior (truth tellers, but not liars, displaying stereotypical lie behavior) on their deception detection accuracy. Our introduced boost involves relying on a single cue, that is, the level of detail. To rule out that its effectiveness is due to simply relying on a single cue—no matter which one—we added a second control condition in Study 3, where people were asked to rely on a nondiagnostic cue (eye gaze aversion).

4 | Study 1

Study 1 examined the performance of the detailedness heuristic in lie–truth discrimination for text versus video statements. After a pilot study ($n = 41$; a detailed report of the pilot can be found in the online supplement), we preregistered Study 1 to test the following hypotheses:

Hypothesis 1. *Deception detection accuracy is better for participants relying on the detailedness heuristic than for participants making unguided veracity judgments, using any cue they like.*

Hypothesis 2. *Deception detection accuracy is better when evaluating text than video statements.*

Hypothesis 3. *The benefit of the detailedness heuristic as compared to the unguided deception judgments is more pronounced for text statements than for video statements.*

5 | Method

The study was approved by the ethics committee of the University of Amsterdam (approval number: 2021-CP-13974). The hypotheses, analyses, and experimental procedure have been preregistered here: https://aspredicted.org/QP8_XFY. Ethics approval, analysis scripts, the data, the experiment script, a pdf document of the survey, and all supplementary information can be found in the online supplement: https://osf.io/7wtyz/. To protect participants' privacy who are featured in the video statements, the statements that we used as stimulus material can be obtained from the corresponding author after signing a nondisclosure agreement.

The study had a 2 (judgment-type: detailedness heuristic vs. unguided judgments) $\times 2$ (medium: text vs. video) $\times 2$ (statement ground-truth: true vs. false) mixed-subjects design, with judgment-type and medium as between-subjects factors and statement veracity as within-subjects-factor. Participants saw 12 video or text statements (half deceptive, half truthful) and either directly made unguided veracity judgments (unguided judgment condition), or first a detailedness judgment and used this for their veracity judgment (in the detailedness heuristic condition). Note that statement veracity was a factor in the design, but not in the analyses, because our hypotheses concerned accuracy in lie–truth discrimination (see online supplement for exploratory analyses testing for within-subjects effect of statement veracity).

5.1 | Sample

We aimed to collect data from 110 participants (assuming exclusion of about 10 participants) by drawing on previous literature investigating when to expect stable and reliable results in lie detection research. Levine et al. (2021) showed in a simulation study that lie detection experiments should have 12 senders, as well as 100 judges per condition to detect meta-analytically found results. Assuming this, our proposed sample size of 12 statements $\times 100$ judges/2 conditions $= 600$ total judgments is lower than the recommendation by Levine et al. (2021). However, they showed that already 500 judgments might yield reasonably reliable results. For budget reasons (full cost for current sample size was £824.59), we therefore chose to stick to $N \sim 110$.

113 participants, recruited via Prolific (2022), started our experiment and were compensated with 5.63 pounds. We only recruited fluent English speakers who did not participate in studies 2, 3, 4, 5, or 7 conducted by Verschuere et al. (2023). Two participants were excluded because they did not consent to the study procedure. As preregistered, participants who failed the two attention checks were excluded from the analyses. This was true for two participants, leaving us with a final sample size of 109 participants, see Table 1 for demographics (obtained from Prolific) and self-report measures. Exploratory analyses of differences between conditions can be found in the online supplement.

5.2 | Procedure

Participants accessed the experiment, gave their informed consent to the study procedure, were randomly assigned to one of

TABLE 1 | Study 1 sample information.

	Text, detailedness	Text, unguided	Video, detailedness	Video, unguided	Overall
N	28	28	24	29	109
M age (*SD*)	24.00 (4.06)	28.18 (6.60)	27.92 (7.83)	24.24 (4.31)	26.00 (6.07)
Percentage female	57.14%	71.43%	66.67%	75.86%	67.89%
M motivation (*SD*)	7.85 (1.56)	7.04 (2.84)	7.17 (1.43)	8.62 (1.63)	7.69 (2.04)
M difficulty (*SD*)	6.25 (2.08)	7.04 (1.86)	7.13 (1.69)	6.72 (2.72)	6.77 (2.14)
Cue "details" mentioned (percentage of participants that mentioned the cue)	18 (66.67%)	15 (53.57%)	19 (79.17%)	19 (70.37%)	71 (66.98%)
Nonverbal cues mentioned (percentage of participants that mentioned the cue)	0 (0%)	1 (3.57%)	10 (41.67%)	18 (66.67%)	29 (27.36%)

the four experimental conditions, and instructed about their tasks during the study. After answering questions to audio- and attention-check questions, participants completed and reviewed the 12 text- or video statements (between-subjects) and made either unguided veracity judgments (unguided judgment condition) or used their heuristic detailedness judgments for the veracity evaluation (detailedness heuristic condition; between-subjects). Statements were presented in random order. Finally, they answered two questions about their motivation and the perceived difficulty of the tasks, were debriefed, and thanked.

5.2.1 | Task Instructions

Participants in the unguided judgment condition were asked to provide a binary veracity judgment: "You will rate whether the person giving the statement is telling a lie or a truth." Participants in the detailedness heuristic condition were first instructed to rate a statement's detailedness (from *0 = not detailed at all*, to *10 = very detailed*) following DePaulo et al.'s (2003) definition (i.e., "Detailedness is the degree to which the message includes details such as descriptions of people, places, actions, objects, events, and the timing of events; degree to which the message seemed complete, concrete, striking, or rich in details"). Subsequently, they received the following decision rule for their binary veracity judgment: "Based on your detailedness judgment, you make a decision of whether the person is telling a truth or a lie, following this rule: Please use your own judgment of detailedness to make a decision, whether the person is telling a truth, or a lie: If you judged detailedness to be equal to or above 6, please select that the person is telling a truth, otherwise (5 or below) select that the person is telling a lie." Participants applied this rule in 94.71% of the cases. The cutoff value of six was an informed guess, splitting the scale in half. In Study 6, Verschuere et al. (2023) showed that this cutoff leads to good truth–lie classifications (see the online supplement for results using different cutoffs). Participants' accuracy was incentivized by informing them about a lottery randomly selecting three of the best-performing individuals that successfully passed all attention checks for a bonus payment of five pounds.

5.2.2 | Audio and Attention Checks

To ensure that participants in the video condition could hear the contents of the videos properly, we showed them two audio checks. Failure of these audio checks would have led to exclusion, which did not occur. Participants in all conditions then saw an example statement to make them acquainted with the setup. This example statement was an additional statement that came from the same set as the rest of the stimuli but was not used in analyses. Participants were asked to perform the same task as for the upcoming stimuli. The example statement was followed by an attention check that asked about the content of the statement (not used as exclusion criterion). Participants were informed that they would be randomly asked attention-check questions throughout the experiment and that only participants that passed all were eligible for the 5 pounds bonus-payment. Attention check questions can be found in the online supplement. The example statement also served the purpose to check whether participants in the detailedness heuristic condition correctly applied the decision rule. If this was not the case, we re-iterated their task before proceeding with the actual experiment.

5.2.3 | Motivation, Difficulty, Cues Used and Debriefing

After completing the experiment, all participants stated their motivation ("How motivated were you to accurately judge the statements?" from 0 = *not motivated at all* to 10 = *very motivated*), as well as perceived difficulty of judging the statements ("How difficult did you find it to accurately judge the statements?" from 0 = *not difficult at all*, to 10 = *very difficult*; see Table 1). Additionally, participants answered an open question, asking them to mention cues they used for making their deception judgments. Descriptive results of motivation, difficulty, and cues mentioned can be found online. Demographic information about their age, gender, and native language was retrieved from Prolific. Finally, they were debriefed and redirected to Prolific.

5.3 | Statements

We used 12 statements (half truthful, half deceptive). These statements were obtained in a mock crime study (Study 6; Verschuere et al. 2023). Participants in this study were undergraduate students from the University of Amsterdam instructed to find a study location (truth condition) or steal a copy of an exam (lie condition). Afterwards, all participants claimed to be innocent of the accused crime (having stolen the exam) and should state they searched for a study location to a condition-blind experimenter while being videotaped. The whole set of stimuli consists of 44 statements, six of which had to be excluded because participants did not give consent for further use (four), a face mask was worn (one), or the task was not completed (one). We selected 12 out of the remaining 38 statements according to the following criteria: To keep the study at a reasonable length, we only selected statements shorter than 4 min. From this subset, we selected statements to ensure that truthful and deceitful statements contain the same number of male and female senders (one male each) and the same number of native vs. non-native speakers (two non-natives each). Since truthful and deceitful statements differ a priori in length and number of details, we did not use this as a selection criterion but analyzed post hoc if there were differences between the statements (the same was true for gaze aversion, see below). The text statements used in our study are the transcribed video statements. Therefore, we used the same 12 statements in both the video and text statement condition.

As a manipulation check of detailedness as a cue to deception, we assessed the statements on their level of detailedness in two ways, both of which indicate that the truthful statements were indeed more detailed than the deceptive ones. First, the original, condition-blind interviewers judged detailedness at the end of the interview (from 0 = *not detailed at all* to 10 = *very detailed*) (see Study 6; Verschuere et al. 2023). Investigating the whole set of 44 statements on this measure shows that truthful statements contained more details ($M = 7.17$, $SD = 1.34$) than deceitful statements ($M = 4.48$, $SD = 1.57$), $BF_{10} = 47028.67$. This also held for the subset of 12 statements that we selected in the current study ($M_{truth} = 6.17$, $SD = 1.17$ versus $M_{deceitful} = 4.17$, $SD = 1.17$; $BF_{10} = 4.11$; Cohen's $d = 1.7$). Second, we used the Linguistic Inquiry and Word Count software (LIWC, Pennebaker et al. 2015) to count the number of perceptual and contextual details (Schutte et al. 2021). Across the entire set, truthful statements ($M = 136.3$, $SD = 52.82$) included more perceptual and contextual details than deceitful statements ($M = 80.72$, $SD = 41.67$), $BF_{10} = 67.41$. In the subset of 12 videos we used, we only found anecdotal support for the assumption that truthful statements ($M = 102.32$, $SD = 26.32$) have more details than false statements ($M = 72.35$, $SD = 32.51$), $BF_{10} = 1.14$.

6 | Deviations From Preregistration

The preregistration stated that the attention check questions would have four answer options, but they had five answer options. Apart from that, we did not deviate from the preregistration.

7 | Results

All analyses were conducted using *RStudio* (Version 1.3.1093; RStudio Team 2020) and *R* (Version 4.0.4; R Core Team 2019).

We ran Bayesian analyses using the *BayesFactor*-package (Version 0.9.12–4.3; Morey et al. 2015) with its default settings. This package uses separate noninformative Jeffreys priors (or JZS-priors) for the effects in an ANOVA equation (Rouder et al. 2012), which is considered to be objective, therefore making as little assumptions about the range of expected effect sizes (Rouder et al. 2009). Robustness checks of our analyses, using different (informed) priors, as well as frequentist analyses of our main analyses can be found in the online supplement: https://osf.io/7wtyz/. Resulting Bayes factors indicate evidence in favor of the alternative ($BF_{10} > 1$) or the null model ($BF_{01} > 1$) (Jarosz and Wiley 2014). Common criteria for interpreting Bayes factors consider a Bayes factor (BF_{10}) close to 1 as inconclusive, between 1 and 3 as weak, between 3 and 10 as moderate, between 10 and 30 as strong, between 30 and 100 as very strong, and above 100 as extreme evidence (Jarosz and Wiley 2014; Wetzels et al. 2011).

7.1 | Preregistered Analyses

We used deception detection accuracy (i.e., the percentage of correct truth–lie classifications) as the dependent variable. We investigated whether results vary with statement veracity, which can be found online. We used a Bayesian 2 (judgment-type: detailedness heuristic vs. unguided judgments) × 2 (medium: text vs. video) between-subjects ANOVA to evaluate our hypotheses. The analysis yields a Bayes factor for the two main effects, as well as one for the interaction effect, comparing the likelihood of the data given the null hypothesis with the likelihood of the data given the alternative hypothesis. Therefore, the Bayes factor tells us how much more likely the data are under the alternative hypothesis compared to the null hypothesis (i.e., the null model).

As predicted, participants' deception detection accuracy using the detailedness heuristic ($M = 60.9\%$, $SD = 15.7\%$) was better than participants' accuracy in making unguided judgments ($M = 52.63\%$, $SD = 14.45\%$), Cohen's $d = 0.55$ (95% CI [0.16, 0.94]). The Bayes factor of $BF_{10} = 7.36$ indicates that the data are 7.36 times more likely to occur under a model that included the main effect for judgment-type than under the null model.

We found no support for Hypothesis 2 that participants' deception detection accuracy would be higher for text statements ($M = 58.18\%$, $SD = 13.72\%$) than for video statements ($M = 55.06\%$, $SD = 17.09\%$), $d = 0.20$ (95% CI [−0.18, 0.58]), $BF_{10} = 0.33$. As it becomes clear from Figure 1, the data also did not point to an interaction effect, $BF_{10} = 0.8$. Thus, the data were equally likely under the null and the alternative model.

7.2 | Study 2

Study 1 replicated the finding that people relying only on detailedness perform better in deception detection than people making unguided judgements, free to use any cue they like.

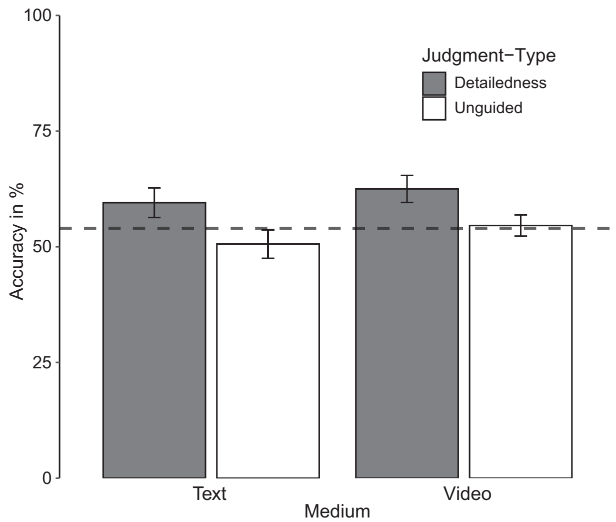

FIGURE 1 | Deception Detection Accuracy (in %) for Text versus Video in the Detailedness Heuristic vs. the Unguided Judgment-Condition. *Note.* On the *y*-axis, you can see the mean deception detection accuracy for the four conditions. On the *x*-axis, the medium conditions can be seen, and the shading of the bars indicates judgment-type condition. The dashed line shows the meta-analytically found mean accuracy of lie detection (54%; DePaulo et al. 2003). The error bars show the standard errors of the mean.

We found no evidence that having access to nonverbal behavior harmed the detailedness heuristic. At the same time, the Bayesian analysis showed that the evidence was not conclusive, possibly as a result of the modest sample size.

For Study 2, we strengthened the experimental manipulation. We created a set of statements in which nonverbal information pointed judges the wrong way. Using a universal and strong stereotypical cue to deception, this biased set contained videos of truth tellers diverting their eye gaze and liars looking the interviewer in the eye. In the control set, eye gaze did not differ between liars and truth tellers. While Study 1 showed that the detailedness heuristic may be resistant to the mere presence of nonverbal behavior, Study 2 allowed us to investigate the potential biasing impact of strongly stereotypical nonverbal behavior. We preregistered Study 2 to test the following hypotheses:

Hypothesis 1. *Participants using the Take-the-best[1]heuristic outperform participants making unguided veracity judgments in deception detection accuracy.*

Hypothesis 2. *Compared to the unguided judgements, deception detection accuracy using the Take-the-best heuristic will be deteriorated less by the presence of biasing nonverbal information.*

8 | Method

The study was approved by the ethics committee of the University of Amsterdam (approval number: FMG-2376). The hypotheses, analyses, and experimental procedure have been preregistered here: https://aspredicted.org/PZ5_VS8. Ethics approval, analysis scripts, the data, the experiment script, a pdf-document of the survey, and supplementary information can be found in the online supplement.

The study had the same 2 (judgment type: detailedness heuristic vs. unguided judgments) × 2 (bias: present vs. absent) × 2 (statement ground truth: true vs. false) mixed-subjects design as Study 1, with judgment type and bias as between-subjects factors and statement veracity a within-subjects-factor. Participants saw 8 video statements (half deceptive, half truthful) and—as in Study 1—either directly evaluated their veracity (unguided judgments condition), or their detailedness and used that for their veracity judgment (in the detailedness heuristic condition). The study procedure and materials used in Study 2 were the same as in Study 1, with the following key differences: (1) Participants now evaluated statements that were either biased or unbiased (between subjects), (2) to reduce the burden of participants, videos were cut in length and participants reviewed 8 instead of 12 videos (3), and we used a Bayesian sequential stopping rule. In the following, we will only describe differences to Study 1.

8.1 | Sample

We adopted a preregistered Bayesian sequential stopping rule (Schönbrodt et al. 2017; Stefan et al. 2019), evaluating the evidence after each month of data collection and ending data collection when the data allowed to support ($BF_{10} > 5$) or refute ($BF_{01} > 5$) our hypotheses. Data collection was stopped after 1 month because the data allowed to endorse Hypothesis 1 and refute Hypothesis 2.

89 participants completed the survey via https://lab.uva.nl, a research portal of the University of Amsterdam, and were compensated with course credits. One participant was excluded because they did not consent to the study procedure. This international undergraduate sample ($n = 88$) was 78.4% female ($M_{age} = 20.5$, $SD_{age} = 3.2$), 19.3% male ($M_{age} = 21.1$, $SD_{age} = 3.0$) with 2.3% preferring not to share their gender. Further descriptives are presented in Table 2.

8.2 | Statements

As in Study 1, we selected statements from the mock crime study by Verschuere et al. (2023); Study 6. To reduce the burden on our participants, we cut the videos, only retaining the first 1/3ed of the video. This cutting procedure preserved the length difference of the videos. We created two sets of 8 statements (4 lies, 4 truths) each. For the biased set, we selected truth tellers that tended to avoid looking at the interviewer and liars that tended to look at the interviewer. For the control set, we selected videos where liars and truth tellers did not differ in their eye gaze behavior. Eye gaze behavior was assessed by a condition-blind researcher (NC) using a stopwatch. We calculated the proportion of time that the interviewee looked at the interviewer. Detailedness was assessed in the same ways explained in detail in Study 1. Table 3 describes the two sets of statements. In both sets, truths were more detailed than lies, with the two sets differing in the presence of strongly stereotypical behavior.

TABLE 2 | Study 2 sample information.

	Bias, detailedness	Bias, unguided	No bias, detailedness	No bias, unguided	Overall
N	21	24	22	21	88
M age (*SD*)	20.57 (2.11)	19.62 (1.28)	20.68 (3.12)	21.81 (4.87)	20.64 (3.14)
Percentage female	71.43%	87.50%	81.82%	71.43%	78.41%
M motivation (*SD*)	6.86 (1.98)	7.29 (2.20)	7.55 (1.60)	7.05 (1.60)	7.19 (1.86)
M difficulty (*SD*)	5.95 (2.09)	7.92 (1.53)	6.00 (1.77)	7.67 (1.56)	6.91 (1.95)
M Reliance on cue "detailedness" (*SD*)	7.30 (2.15)	7.04 (2.22)	9.11 (1.18)	6.80 (1.85)	7.50 (2.09)
M Reliance on cue "eye gaze aversion" (*SD*)	4.35 (2.41)	5.12 (2.72)	4.78 (2.08)	6.30 (2.30)	5.15 (2.47)
M Reliance on cue "emotions" (*SD*)	5.85 (2.56)	5.96 (2.37)	5.17 (2.68)	5.90 (2.10)	5.74 (2.40)

Note: The statistics for the cues stem from a question asking participants how much they used the respective cue for their judgments on a scale from 0 = *not at all*, to 10 = *very much*. Due to a technical error, for some participants, there were missing data recorded for the cue reliance. Specifically, no data on these variables are present for four participants in the no bias, detailedness condition, and for one person each in the bias detailedness and the no bias unguided judgment condition.

TABLE 3 | Characteristics of the two sets of statements (biased versus non-biased).

	Bias present		Bias absent	
	Truthful	**Deceptive**	**Truthful**	**Deceptive**
Proportion eye gaze aversion	M = 80% SD = 7%	M = 29% SD = 3%	M = 56% SD = 5%	M = 55% SD = 13%
Statement length (LIWC Word Count)	M = 717.00 SD = 318.96	M = 469.75 SD = 255.43	M = 693.50 SD = 121.86	M = 262.00 SD = 42.17
Detailedness judgement (0–10) by original interviewer	M = 6.50 SD = 0.57	M = 4.50 SD = 2.65	M = 7.00 SD = 1.63	M = 3.00 SD = 0.81
LIWC count of perceptual and contextual details	M = 120.99 SD = 50.21	M = 72.51 SD = 36.79	M = 133.97 SD = 17.80	M = 40.75 SD = 6.64
Number of verifiable details	M = 8.25 SD = 2.05	M = 5.00 SD = 5.40	M = 11.50 SD = 7.95	M = 4.50 SD = 2.61

Note: Number of verifiable details was evaluated for the entire video in an earlier study.

9 | Deviations From Preregistration

We planned to exclude participants who failed the two attention checks. Due to a programming error, the attention check was only included in two of the four conditions and using it would imply condition-dependent exclusions. We therefore chose not to exclude anyone. Note that in the conditions where we had recorded the attention check, no participants failed both checks.

10 | Results

10.1 | Preregistered Analyses

All analyses were conducted using *RStudio* (Version 1.3.1093; RStudio Team 2020) and *R* (Version 4.0.4; R Core Team 2019). Frequentist analyses can be found in the online supplement.

We ran a Bayesian 2 × 2 ANOVA with judgment type (judgement type: detailedness heuristic vs. unguided judgments) and bias (bias: present vs. absent) as between-subjects factors and percentage of correct classifications as the dependent variable. We found strong evidence for a main effect of judgement method, indicating that participants' deception detection accuracy was higher when relying on the detailedness heuristic (M = 71.8%, SD = 19.3%) compared to making unguided veracity judgements (M = 56.9%, SD = 20.4%), Cohen's d = 0.76 (95% CI [0.32, 1.19]), BF_{10} = 39.57.

However, as Figure 2 clearly illustrates, the advantage of the heuristic method over intuitive judgments depended on the presence of biasing nonverbal behavior. Comparing the model that included the interaction effect with the null model that included only the intercept showed overwhelming support for an interaction effect, BF_{10} = 4195.67. Thus, in contrast to our predictions, deception detection accuracy using the detailedness heuristic was affected more by the presence of biasing nonverbal information than the unguided judgements.

10.2 | Non-Preregistered Analyses

Breaking down the interaction, we conducted follow-up t-tests for the bias present and bias absent conditions, separately. In the bias-absent conditions, the detailedness heuristic (M = 83.5%,

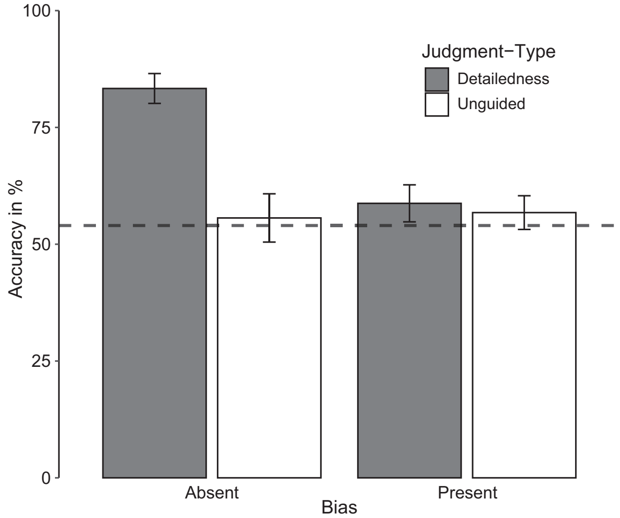

FIGURE 2 | Deception Detection Accuracy (in %) in the Detailedness Heuristic versus the Unguided Judgment Condition in the Presence and Absence of Biasing Nonverbal Behavior. *Note.* On the *y*-axis, you can see the mean deception detection accuracy for the four conditions. On the *x*-axis, the bias conditions can be seen, and the shading of the bars indicates judgment-type condition. The dashed line shows the meta-analytically found mean accuracy of lie detection (54%; DePaulo et al. 2003). The error bars show the standard errors of the mean.

$SD = 12.4\%$) clearly outperformed unguided judgements ($M = 57.1\%$, $SD = 23.6\%$), Cohen's $d = 1.44$ (95% CI [0.76, 2.11]), $BF_{10} = 506.24$. But in the face of bias, the detailedness heuristic ($M = 59.5\%$, $SD = 17.6\%$) did no longer outperform unguided judgements ($M = 56.8\%$, $SD = 17.7\%$), Cohen's $d = 0.16$ (95% CI [−0.43, 0.74]), $BF_{10} = 0.33$.

Finally, participants who were in the detailedness heuristic condition applied the decision rule 97.09% of the time.

10.3 | Study 3

Study 2 showed that the increase in deception detection accuracy when making heuristic-based judgments completely disappeared in the face of bias (i.e., when truth tellers avoided gaze but liars did not). This finding is surprising, given that participants were instructed to only use a statement's level of detailedness for deception detection, which should, thanks to its higher diagnostic value, allow for a better discrimination between true and false statements. Because this finding was unexpected, we replicated the study. With Study 3, we aimed to answer three questions. First, we wanted to test whether participants use the cue they are told to use when evaluating statements. To this end, we introduced a third condition where participants were instructed to use a nondiagnostic cue for making veracity judgments: The amount of gaze aversion. If participants followed our instructions and did use this cue in their judgments, they should perform worse on the biased (i.e., truth tellers avoided eye gaze, liars did not) than on the nonbiased video statements.

Hypothesis 1. *As a manipulation check, people judging deception based on eye gaze aversion should perform worse on the biased set than on the control set.*

Second, we aimed to replicate the finding from Studies 1 and 2 that participants making detailedness-based judgments achieve a higher deception detection accuracy than participants making unguided judgments. Yet, it remains unclear if this increased accuracy is due to participants making judgments based on a more diagnostic cue (i.e., a statement's level of detailedness) or because they were simply more engaged or attentive than participants who "only" judged a statement's veracity. If the former is true, participants using another, nondiagnostic cue for making heuristic judgments (i.e., eye gaze aversion) should not perform better than participants making unguided judgments.

Hypothesis 2. *Judging detailedness will lead to higher deception detection accuracy compared to both unguided judgements as well as judgements based on eye gaze.*

Finally, we aimed to increase power to obtain conclusive evidence in favor or against the surprising finding from Study 2 that participants making detailedness-based judgments achieved the same deception detection accuracy as participants making unguided judgments. In line with our predictions in Studies 1 and 2, but at odds with the surprising finding from Study 2, we formulated hypothesis 3 as follows.

Hypothesis 3. *Even in the face of bias, judgements based on detailedness will be more accurate than unguided judgements.*

11 | Method

The study was approved by the ethics committee of the University of Amsterdam (approval number: 2021-CP-13171). The hypotheses, analyses, and experimental procedure have been preregistered here: https://aspredicted.org/tn89r.pdf. Ethics approval, analysis scripts, the data, and supplementary information can be found in the online supplement.

Study 3 had a 3 (judgment-type: detailedness heuristic vs. eye gaze heuristic vs. unguided judgments) ×2 (bias: present vs. absent) ×2 (statement ground-truth: true vs. false) mixed-subjects design, with judgment type and bias as between-subjects factors and statement veracity a within-subjects factor. Participants saw 8 video statements (half deceptive, half truthful) and—as in Study 1 and 2—either directly evaluated their veracity (unguided judgment condition), or their detailedness and used that for their veracity judgment (in the detailedness heuristic condition). Participants in the third judgment-type condition (i.e., eye gaze heuristic) were instructed to evaluate the amount the statement giver averted their eye gaze and use that for their veracity judgment. The study procedure and materials used in Study 3 were the same as in Study 2, with the following key change. We introduced a third level in the judgment condition, that is, the eye gaze heuristic condition. In the following, we will only describe differences to Study 2.

11.1 | Sample

We adopted the same preregistered Bayesian sequential stopping rule as in Study 2. Data collection ran from February 7th, 2024 to March 7th, 2024 and was stopped thereafter because the data allowed to support all three hypotheses.

149 participants, all students from the University of Amsterdam, participated in the survey via https://www.lab.uva.nl and were compensated with course credits. As preregistered, we excluded five participants who failed both attention checks. Out of the remaining participants, one participant did not finish the study and was consequently excluded from the analyses, which left us with a total sample size of 143 participants.[2] Out of these, 104 (72.70%) self-identified as female ($M_{age}=20.6$, $SD_{age}=3.46$), 37 (25.90%) as male ($M_{age}=20.9$, $SD_{age}=3.96$) and one participant indicated to identify differently (age = 22). Further descriptive statistics are presented in Table 4.

11.1.1 | Task Instructions

Participants in the eye gaze heuristic condition received a definition of eye gaze aversion (i.e., "Eye gaze aversion refers to the behavior of actively avoiding or limiting direct eye contact with others, in this case the interviewer") and were instructed to rate the amount of eye gaze aversion (from 0 = *averts eye gaze all the time*, to 10 = *no eye gaze aversion at all*): "Please rate the following statement on its level of eye gaze aversion." Participants received the same decision rule as in the detailedness heuristic condition; however, using their eye gaze judgment: "Is the person telling a lie or a truth? If you judged eye gaze aversion to be equal to or above 6, choose 'telling a truth', otherwise choose 'telling a lie'." Participants applied this rule 95.70% of the time. Participants in the detailedness heuristic condition applied the decision rule 92.40% of the time.

11.1.2 | Cues Used and Debriefing

After completing the experiment, participants also indicated, on a scale from 0 = *not at all* to 10 = *very much*, to what degree they used one of five cues for making their veracity judgments: "Which cue did you rely on when judging the video statements?". The five cues we asked about where detailedness, eye gaze aversion, emotionality, consistency, and fidgeting. You can find an overview and analyses regarding these cues in the online supplement. Due to a technical error,[3] 10 participants did not record an answer to these questions and were excluded from all analyses involving this measurement. After that, participants were debriefed and thanked.

12 | Deviations From Preregistration

Other than preregistered and to improve reproducibility, we ran all reported analyses in R, instead of JASP (which is based on R). Results were the same for both software.

13 | Results

13.1 | Preregistered Analyses

All analyses were conducted using *RStudio* (Version 2023.6.2.561; RStudio Team 2020) and *R* (Version 4.3.2; R Core Team 2019). Frequentist analyses can be found in the online supplement.

To test Hypothesis 1, we conducted a Bayesian independent sample *t*-test comparing the deception detection accuracy of participants in the eye gaze heuristic condition judging biased videos with the same accuracy of participants in the eye gaze heuristic condition judging unbiased videos. We found extreme evidence in favor of the hypothesis that when making heuristic-based judgments using eye gaze as a cue to deception, participants judging biased videos performed worse ($M=19\%$, $SD=19.5\%$) than participants judging unbiased videos ($M=52.2\%$, $SD=15.8\%$), Cohen's $d=1.91$ (95% CI [1.2, 2.6]), $BF_{10}=93091.3$. Note that the accuracy achieved using eye gaze as a cue for detecting deception on unbiased (i.e., "realistic") videos (using a one-sample *t*-test) did not allow participants to perform better than chance, $BF_{10}=0.266$.

For testing Hypothesis 2, we conducted a 3 (judgment type: detailedness heuristic vs. unguided judgments) × 2 (bias: present vs. absent) between-subjects Bayesian ANOVA. We found extreme evidence for a main effect of judgment type in favor of Hypothesis 2. Participants making heuristic judgments using detailedness as a cue achieved a higher deception detection accuracy ($M=73.4\%$, $SD=17.8\%$) than participants making heuristic judgments using eye gaze aversion ($M=35.6\%$, $SD=24.3\%$) and participants making unguided judgments ($M=48.2\%$, $SD=18.6\%$), $BF_{10}>2.024*10^{20}$.

TABLE 4 | Study 3 Sample information.

	Bias, D	Bias, E	Bias, U	No bias, D	Bias, E	No bias, U	Overall
N	25	23	25	23	23	23	142[†]
M age (*SD*)	20.7 (4.29)	21.4 (4.23)	20.6 (3.32)	20.7 (2.22)	21 (4.73)	19.6 (1.62)	20.7 (3.57)
Percentage female	68%	91.3%	60%	78.3%	69.6%	73.9%	73.2%
M motivation (*SD*)	8.12 (1.64)	7.65 (1.94)	7.68 (1.18)	7.52 (1.97)	8.04 (1.52)	7.65 (1.23)	7.78 (1.59)
M difficulty (*SD*)	5.52 (2.02)	6.43 (2.23)	8.2 (1.44)	6 (1.83)	6.39 (1.41)	7.65 (1.8)	6.7 (2.02)

Note: The statistics for the cues stem from a question asking participants how much they used the respective cue for their judgments on a scale from 0 = *not at all*, to 10 = *very much*.
Abbreviations: D, detailedness heuristic condition; E, eye gaze heuristic condition; U, unguided judgment condition.
[†]This number is different from the actual sample size of $N=143$, as one participant did not provide sociodemographic information.

To test Hypothesis 3, we conducted a Bayesian independent sample *t*-test comparing the deception detection accuracy of participants in the detailedness heuristic condition with the accuracy of participants in the unguided judgment condition, both judging biased videos. We found extreme evidence in favor of the hypothesis that—when evaluating biased videos—participants making detailedness heuristic-based judgments achieved a higher accuracy ($M = 68\%$, $SD = 17.3\%$) than participants making unguided judgments ($M = 39.4\%$, $SD = 14\%$), Cohen's $d = 1.85$ (95% CI [1.19, 2.51]), $BF_{10} = 239{,}922$. The results, displaying mean accuracy for all conditions, can be seen in Figure 3.

13.2 | Nonpreregistered Analyses

In Study 2, we found that the advantage of making detailedness heuristic-based deception judgments disappeared in the face of bias. The results of testing Hypothesis 3 (see above) show that this is not the case—participants making such judgments achieved a higher accuracy than participants making unguided judgments. But are judgments using the detailedness heuristic "immune" to bias? To test this, we conducted a Bayesian independent sample *t*-test comparing the accuracy of participants from the two groups making deception detection judgments using the detailedness heuristic. We found that the detailedness heuristic cannot fully counter the harmful effect of bias. Participants making detailedness heuristic-based judgments for biased videos performed worse ($M = 68\%$, $SD = 17.3\%$) than participants making the same judgments for unbiased videos ($M = 79.3\%$, $SD = 16.7\%$), Cohen's $d = 0.68$ (95% CI [0.09, 1.26]), $BF_{10} = 2.368$.

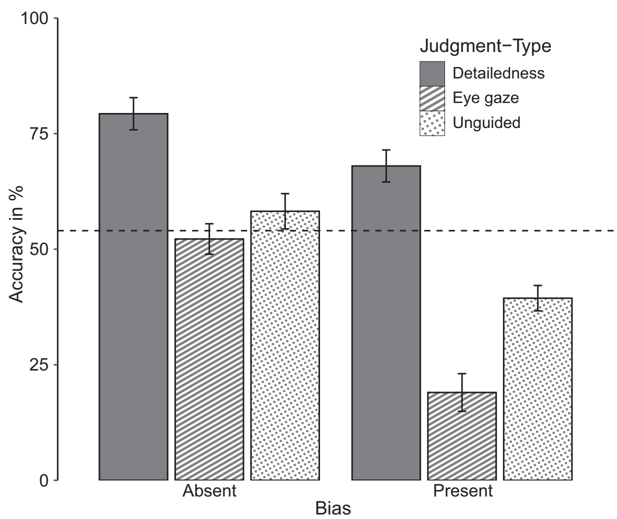

FIGURE 3 | Deception Detection Accuracy (in %) in the Detailedness Heuristic versus Eye Gaze Heuristic versus Unguided Judgment-Condition in the Presence and Absence of Biasing Nonverbal Behavior. *Note.* On the *y*-axis, you can see the mean deception detection accuracy for the four conditions. On the *x*-axis, the bias conditions can be seen, and the shading of the bars indicate judgment-type conditions. The dashed line shows the meta-analytically found mean accuracy of lie detection (54%; DePaulo et al. 2003). The error bars show the standard errors of the mean.

13.2.1 | Cue Usage

We analyzed participants' self-reported cue usage. As expected, we found that participants in the detailedness heuristic condition relied more strongly on detailedness than participants in either of the two other conditions. Participants in the eye gaze heuristic condition, on the other hand, relied strongly on eye gaze aversion when making their judgments and more importantly, more than participants in the unguided judgment or the detailedness heuristic condition. Detailed analyses can be found in the online supplement.

14 | Discussion

With the present studies, we investigated whether people's accuracy in lie detection can be improved by giving them an easy-to-use heuristic decision rule and whether potential improvements depend on the presence of (biasing) nonverbal behavior. A simple heuristic, specifically, to rely only on the level of detail in a given statement, led to higher deception detection accuracy than unguided deception judgments in all three studies. We found no impact of merely having access to nonverbal behavior (i.e., video vs. text). Strongly biasing nonverbal information nullified the benefit of heuristic judgements over unguided judgements in Study 2; however, there was only weak to moderate evidence for the absence of a difference ($BF_{10} = 0.33$). This was not the case in Study 3. Although we again found evidence that heuristic judgements were not fully immune to strongly biasing nonverbal behavior, we also found conclusive evidence ($BF_{10} = 239{,}922$) that participants making heuristic-based judgments achieved a higher accuracy in detecting deceit in the face of bias compared to participants making unguided judgements. This difference was due to an advantage of using a diagnostic cue (i.e., detailedness), as participants using a nondiagnostic cue (i.e., eye gaze aversion) performed at (nonbiased videos) and below (biased videos) chance level.

14.1 | The Heuristic Approach

We found that participants making unguided judgements, despite being free to use any cue they like, performed close to the chance level. This fits with a large body of evidence that people are poor lie detectors (Bond and DePaulo 2006). In contrast to this, we found that a simple rule of thumb—to rely only on the level of detail in the message—facilitated deception detection. While this is certainly counterintuitive, the less-is-more phenomenon fits with decision-making theory and evidence from various research fields (Fiedler and Walka 1993; Gigerenzer and Gaissmaier 2011; Gigerenzer and Goldstein 1996; Verschuere et al. 2023). This intervention can be considered a "boost", that is, an intervention tailored at improving people's ability to make better decisions (Hertwig and Grüne-Yanoff 2017; Lorenz-Spreen et al. 2021). Deception detection appears to be a domain (see also, Verschuere et al. 2023) where people might be able to relatively easily boost their ability to detect lies by making heuristic-based judgments. Importantly, using a nondiagnostic cue for making heuristic-based judgments (i.e., eye gaze aversion) did not improve a person's accuracy. This clearly illustrates that the heuristic approach is advantageous because of its reliance on diagnostic information, rather than improving participants' judgments by merely using a single cue and making them more attentive or motivated to detect deceit.

The effect sizes for the detailedness heuristic that we observed here (d = 0.5–1.85), and were observed before (Verschuere et al. 2023), are substantially higher than the meta-analytic average of d = 0.3 for the cue detailedness obtained in studies with trained coders by DePaulo et al. (2003). Research since the DePaulo meta-analysis has revealed some moderating factors that may help to explain this difference in effect size. Most importantly, we used alibi scenarios, where senders talk about a rich, past, discrete event. Such scenarios are likely to contain more details than other scenarios (e.g., about opinions), thus increasing obtained effect sizes (Nahari 2018; Verschuere, Bogaard, and Meijer 2021). Moreover, truth tellers provide more details when the interviewer builds rapport (Gabbert et al. 2021; Matsumoto and Hwang 2018), uses an information-gathering interview style (Meissner et al. 2014, 2017), and encourages to say more (Vrij et al. 2017). Since the statements we used in this study were obtained under these conditions, it is plausible to assume that the higher level of detailedness in the used statements might have caused the larger observed effect sizes.

The heuristic approach relates to, but is different from, approaches referred to as "indirect deception detection" (also, "rapid judgements", "snap judgements", "implicit deception detection," and "unconscious lie detection"). What unites these studies and contrasts them from "direct deception detection" (i.e., unguided judgments) is that participants are asked to judge a characteristic other than, but presumably associated with, deception (Street and Richardson 2015), for example, "Is the person thinking hard?" (Vrij et al. 2001). Some studies found indirect deception to be successful, yet others obtained null findings. Reviews have subsequently concluded that "results appear inconsistent, and it is unclear which mechanisms could be responsible for this advantage" (Sporer and Ulatowska 2021, p. 1), or even that the "meta-analysis suggests that accuracy is actually a little higher with direct than with indirect measures" (Levine and Bond 2014, p. 1961). Finally, Granhag (2006) concluded that conceptual clarity is warranted because indirect deception detection can imply judging verbal, nonverbal, or physiological cues from the sender or the receiver, in an indirect, intuitive, or even unconscious way. We aimed at improving the conceptual clarity.

In a situation of cue dominance, where one cue is much more diagnostic than other cues, it can be beneficial to make a decision solely on that dominant cue and ignore all other cues. We have argued that deception detection may be such a situation, with detailedness being the dominant cue. Our findings support that notion and help to explain why studies relying on cues other than detailedness produced weaker or inconsistent findings (Ulatowski 2018; Vrij et al. 2001). Moreover, they explain why studies relying on judging multiple cues—even when one of the cues was detailedness—again produced weaker or inconsistent findings (Dunbar et al. 2017; Evans and Michael 2014; for a direct test, see studies eight–nine in, Verschuere et al. 2023). It is important, however, to keep in mind that these findings might be idiosyncratic to the experimental paradigm used here, requiring further research with naturalistic, real-world studies to allow for recommendations as to whether people should rely solely on detailedness for making deception–detection judgments.

14.2 | Do Nonverbal Cues Facilitate or Interfere With Lie Detection?

In Study 1, we found no effect of medium (video vs. text) on deception detection accuracy. This indicates that having access to nonverbal behavior likely has a small effect on deception detection accuracy. In line with this, Leach et al. (2016) find that restricting the access to nonverbal cues led to a modest 3–6 percentage points improvement in lie detection accuracy (but see Denault et al. 2017). As prominently demonstrated by the example of Elizabeth Holmes and the findings from Studies 2 and 3, this is different when having access to strongly biased information, where a person's deception detection accuracy is undermined. Although the heuristic approach was not completely immune to this kind of bias, it greatly reduced its impact and allowed participants to perform better than chance. Thus, for the presently investigated context, that is, detecting lies, nonverbal cues seem to have a detrimental effect on a person's lie detection ability and reducing the use of these nonverbal cues improves the ability to detect deceit.

14.3 | Limitations

This work is not without its limitations. First, to assure having ground truth in our studies, we used an instructed lying paradigm. It is therefore vital to expand our work to spontaneous lies. Second, when comparing video to text (Study 1), participants had access to more information (e.g., certain demographics and attractiveness) than nonverbal behavior. Even when comparing biased videos to unbiased videos that were matched on certain demographics (Study 2), there may be variables we did not control for. To allow full control over possible confounding factors, future research could involve avatars or actors to display stereotypical behavior (cf. Vrij and Winkel 1991).

15 | Conclusions

Our three studies show that deception detection in interview settings like the one presented here can be improved by relying only on the level of detail in a message. Yet, even when guided to only evaluate detailedness, strongly stereotypical information seems hard to ignore and can reduce deception detection accuracy. As nonverbal behavior carries little diagnostic information for detecting lies in face-to-face interactions, deception detection may therefore best be carried out only on what is said. However, when time or resources are limited, one might want to rely on frugal, easy-to-use heuristics for catching a liar.

Author Contributions

Aaron Benjamin Lob: data curation, software, methodology, investigation, formal analysis, validation, visualization, writing – original draft, writing – review and editing, resources. **Nisrin Chakir:** investigation, methodology, resources. **Laurine van Munster van Heuven:** investigation, methodology, resources. **Bruno Verschuere:** conceptualization, project administration, resources, supervision, writing – original draft, writing – review and editing.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are openly available in OSF at https://osf.io/7wtyz/.

Endnotes

[1] Take-the-best refers to the same heuristic implemented in the other studies, that is, the detailedness heuristic.

[2] Note that one participant did not consent to the third question of the informed consent (i.e., "I declare that (1) I am 16 years of age or older, (2) I have read and understood the information, (3) I consent to participating in the study and to the usage of the obtained data, (4) I will preserve the right to revoke this consent without notice, (5) I will preserve the right to cease the study at any moment desired by me") which was possible due to a programming error. Since this participant consented to the two other questions, we decided to include them in the final sample. Furthermore, due to a programming error, one participant did not provide sociodemographic information. We decided to include this participant's data in all analyses that did not make use of sociodemographic information.

[3] Qualtrics, per default, displays the "thumb" on continuous scales at the lower end (i.e., 0 = not at all). Thus, participants wanting to indicate not using a cue, might have left the thumb unchanged. This, however, is recorded as a missing answer by Qualtrics. After taking note of this issue, we changed the respective items to be compulsory, forcing participants to actively change the location of the thumb (e.g., move it to 0).

References

Aamodt, M., and H. Custer. 2006. "Who Can Best Catch a Liar? A Meta-Analysis of Individual Differences in Detecting Deception." *Forensic Examiner* 15: 6–11.

Akehurst, L., G. Köhnken, A. Vrij, and R. Bull. 1996. "Lay Persons' and Police Officers' Beliefs Regarding Deceptive Behaviour." *Applied Cognitive Psychology* 10, no. 6: 461–471. https://doi.org/10.1002/(SICI)1099-0720(199612)10:6<461::AID-ACP413>3.0.CO;2-2.

Bogaard, G., E. H. Meijer, A. Vrij, and H. Merckelbach. 2016. "Strong, but Wrong: Lay People's and Police Officers' Beliefs About Verbal and Nonverbal Cues to Deception." *PLoS One* 11, no. 6: e0156615. https://doi.org/10.1371/journal.pone.0156615.

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

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

Burgoon, J. K., J. P. Blair, and R. E. Strom. 2008. "Cognitive Biases and Nonverbal Cue Availability in Detecting Deception." *Human Communication Research* 34, no. 4: 572–599. https://doi.org/10.1111/j.1468-2958.2008.00333.x.

Denault, V., L. M. Jupe, O. Dodier, and N. Rochat. 2017. "To Veil or Not to Veil: Detecting Lies in the Courtroom.A Comment on Leach et al. (2016)." *Psychiatry, Psychology and Law* 24, no. 1: 102–117. https://doi.org/10.1080/13218719.2017.1260619.

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

Dressel, J., and H. Farid. 2018. "The Accuracy, Fairness, and Limits of Predicting Recidivism." *Science Advances* 4, no. 1: eaao5580. https://doi.org/10.1126/sciadv.aao5580.

Dunbar, N. E., M. L. Jensen, L. A. Harvell-Bowman, K. M. Kelley, and J. K. Burgoon. 2017. "The Viability of Using Rapid Judgments as a Method of Deception Detection." *Communication Methods and Measures* 11, no. 2: 121–136. https://doi.org/10.1080/19312458.2016.1276893.

Evans, J. R., and S. W. Michael. 2014. "Detecting Deception in Non-Native English Speakers." *Applied Cognitive Psychology* 28, no. 2: 226–237. https://doi.org/10.1002/acp.2990.

Fiedler, K., and I. Walka. 1993. "Training Lie Detectors to Use Nonverbal Cues Instead of Global Heuristics." *Human Communication Research* 20, no. 2: 199–223. https://doi.org/10.1111/j.1468-2958.1993.tb00321.x.

Gabbert, F., L. Hope, K. Luther, G. Wright, M. Ng, and G. Oxburgh. 2021. "Exploring the Use of Rapport in Professional Information-Gathering Contexts by Systematically Mapping the Evidence Base." *Applied Cognitive Psychology* 35, no. 2: 329–341. https://doi.org/10.1002/acp.3762.

Gigerenzer, G., and W. Gaissmaier. 2011. "Heuristic Decision Making." *Annual Review of Psychology* 62, no. 1: 451–482. https://doi.org/10.1146/annurev-psych-120709-145346.

Gigerenzer, G., and D. G. Goldstein. 1996. "Reasoning the Fast and Frugal Way: Models of Bounded Rationality." *Psychological Review* 103, no. 4: 650–669. https://doi.org/10.1037/0033-295X.103.4.650.

Gigerenzer, G., and P. M. Todd. 1999. *Simple Heuristics That Make Us Smart*. New York: Oxford University Press.

Global Deception Research Team. 2006. "A World of Lies." *Journal of Cross-Cultural Psychology* 37, no. 1: 60–74. https://doi.org/10.1177/0022022105282295.

Granhag, P. 2006. "Rethinking Implicit Lie Detection." *Journal of Credibility Assessment and Witness Psychology* 7: 180–190.

Hartwig, M., and C. F. Bond. 2014. "Lie Detection from Multiple Cues: A Meta-Analysis." *Applied Cognitive Psychology* 28, no. 5: 661–676. https://doi.org/10.1002/acp.3052.

Hertwig, R., and T. Grüne-Yanoff. 2017. "Nudging and Boosting: Steering or Empowering Good Decisions." *Perspectives on Psychological Science* 12, no. 6: 973–986. https://doi.org/10.1177/1745691617702496.

Jarosz, A., and J. Wiley. 2014. "What Are the Odds? A Practical Guide to Computing and Reporting Bayes Factors." *Journal of Problem Solving* 7, no. 1: 2–9. https://doi.org/10.7771/1932-6246.1167.

Johnson, M. K., and C. L. Raye. 1981. "Reality Monitoring." *Psychological Review* 88, no. 1: 67–85. https://doi.org/10.1037/0033-295x.88.1.67.

Kahneman, D., S. P. Slovic, P. Slovic, A. Tversky, and C. U. Press. 1982. *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press.

Leach, A.-M., N. Ammar, D. N. England, L. M. Remigio, B. Kleinberg, and B. J. Verschuere. 2016. "Less Is More? Detecting Lies in Veiled Witnesses." *Law and Human Behavior* 40, no. 4: 401–410. https://doi.org/10.1037/lhb0000189.

Leach, A.-M., V. Talwar, K. Lee, N. Bala, and R. C. L. Lindsay. 2004. "Intuitive Lie Detection of Children's Deception by Law Enforcement Officials and University Students." *Law and Human Behavior* 28, no. 6: 661–685. https://doi.org/10.1007/s10979-004-0793-0.

Levine, T. R., and C. F. Bond. 2014. "Direct and Indirect Measures of Lie Detection Tell the Same Story: A Reply to Ten Brinke, Stimson, and Carney (2014)." *Psychological Science* 25, no. 10: 1960–1961. https://doi.org/10.1177/0956797614536740.

Levine, T. R., Y. Daiku, and J. Masip. 2021. "The Number of Senders and Total Judgments Matter More Than Sample Size in Deception-Detection Experiments." *Perspectives on Psychological Science* 17, no. 1: 191–204. https://doi.org/10.1177/1745691621990369.

Lorenz-Spreen, P., M. Geers, T. Pachur, R. Hertwig, S. Lewandowsky, and S. M. Herzog. 2021. "Boosting people's Ability to Detect Microtargeted Advertising." *Scientific Reports* 11, no. 1: 15541. https://doi.org/10.1038/s41598-021-94796-z.

Luke, T. J. 2019. "Lessons From Pinocchio: Cues to Deception May be Highly Exaggerated." *Perspectives on Psychological Science* 14, no. 4: 646–671. https://doi.org/10.1177/1745691619838258.

Masip, J., S. L. Sporer, E. Garrido, and C. Herrero. 2005. "The Detection of Deception with the Reality Monitoring Approach: A Review of the Empirical Evidence." *Psychology, Crime & Law* 11, no. 1: 99–122. https://doi.org/10.1080/10683160410001726356.

Matsumoto, D., and H. C. Hwang. 2018. "Social Influence in Investigative Interviews: The Effects of Reciprocity." *Applied Cognitive Psychology* 32, no. 2: 163–170. https://doi.org/10.1002/acp.3390.

McShane, J. 2019, March 31. ""She didn't blink": Why Elizabeth Holmes' Body Language Is a Talking Point." *Image*. https://www.image.ie/editorial/she-never-blinks-why-elizabeth-holmes-body-language-is-a-talking-point-146103.

Meissner, C. A., A. D. Redlich, S. W. Michael, et al. 2014. "Accusatorial and Information-Gathering Interrogation Methods and Their Effects on True and False Confessions: A Meta-Analytic Review." *Journal of Experimental Criminology* 10, no. 4: 459–486. https://doi.org/10.1007/s11292-014-9207-6.

Meissner, C. A., F. Surmon-Böhr, S. Oleszkiewicz, and L. J. Alison. 2017. "Developing an Evidence-Based Perspective on Interrogation: A Review of the U.S. government's High-Value Detainee Interrogation Group Research Program." *Psychology, Public Policy, and Law* 23: 438–457. https://doi.org/10.1037/law0000136.

Morey, R. D., J. N. Rouder, and T. Jamil. 2015. "Package 'Bayesfactor' [Computer Software]." https://richarddmorey.github.io/BayesFactor/.

Nahari, G. 2018. "Chapter 14—The Applicability of the Verifiability Approach to the Real World." In *Detecting Concealed Information and Deception*, edited by J. P. Rosenfeld, 329–349. San Diego: Academic Press. https://doi.org/10.1016/B978-0-12-812729-2.00014-8.

Nahari, G., A. Vrij, and R. P. Fisher. 2014. "The Verifiability Approach: Countermeasures Facilitate Its Ability to Discriminate Between Truths and Lies." *Applied Cognitive Psychology* 28, no. 1: 122–128. https://doi.org/10.1002/acp.2974.

Pennebaker, J. W., R. L. Boyd, K. Jordan, and K. Blackburn. 2015. "The Development and Psychometric Properties of LIWC2015." https://repositories.lib.utexas.edu/handle/2152/31333.

Popken, B. 2021. "Elizabeth Holmes Is Known for Her Voice. But Nonverbal Communication Could Be Key at Trial." https://www.nbcnews.com/business/business-news/elizabeth-holmes-silently-charming-jurors-theranos-case-end-mistrial-rcna4155, NBC News.

Prolific. 2022. "Prolific." www.prolific.co (2022).

R Core Team. 2019. *R: A Language and Environment for Statistical Com-Puting*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Rouder, J. N., R. D. Morey, P. L. Speckman, and J. M. Province. 2012. "Default Bayes Factors for ANOVA Designs." *Journal of Mathematical Psychology* 56, no. 5: 356–374. https://doi.org/10.1016/j.jmp.2012.08.001.

Rouder, J. N., P. L. Speckman, D. Sun, R. D. Morey, and G. Iverson. 2009. "Bayesian t Tests for Accepting and Rejecting the Null Hypothesis." *Psychonomic Bulletin & Review* 16, no. 2: 225–237. https://doi.org/10.3758/PBR.16.2.225.

RStudio Team. 2020. "RStudio: Integrated Development Environment for R." RStudio, PBC [Computer software]. http://www.rstudio.com/.

Salganik, M. J., I. Lundberg, A. T. Kindel, et al. 2020. "Measuring the Predictability of Life Outcomes With a Scientific Mass Collaboration." *Proceedings of the National Academy of Sciences* 117, no. 15: 8398–8403. https://doi.org/10.1073/pnas.1915006117.

Schönbrodt, F. D., E.-J. Wagenmakers, M. Zehetleitner, and M. Perugini. 2017. "Sequential Hypothesis Testing With Bayes Factors: Efficiently Testing Mean Differences." *Psychological Methods* 22, no. 2: 322–339. https://doi.org/10.1037/met0000061.

Schutte, M., G. Bogaard, E. M. Giolla, L. Warmelink, B. Kleinberg, and B. Verschuere. 2021. "Man versus Machine: Comparing Manual With LIWC Coding of Perceptual and Contextual Details for Verbal Lie Detection." PsyArXiv. https://doi.org/10.31234/osf.io/cth58.

Sporer, S. L., and B. Schwandt. 2007. "Moderators of Nonverbal Indicators of Deception: A Meta-Analytic Synthesis." *Psychology, Public Policy, and Law* 13, no. 1: 1–34. https://doi.org/10.1037/1076-8971.13.1.1.

Sporer, S. L., and J. Ulatowska. 2021. "Indirect and Unconscious Deception Detection: Too Soon to Give Up?" *Frontiers in Psychology* 12: 601852.

Stefan, A. M., Q. F. Gronau, F. D. Schönbrodt, and E.-J. Wagenmakers. 2019. "A Tutorial on Bayes Factor Design Analysis Using an Informed Prior." *Behavior Research Methods* 51, no. 3: 1042–1058. https://doi.org/10.3758/s13428-018-01189-8.

Steller, M., and G. Köhnken. 1989. "Criteria Based Statement Analysis." In *Psychological Methods in Criminal Investigation and Evidence*, edited by D. C. Raskin, 217–245. New York: Springer.

Street, C. N. H., and D. C. Richardson. 2015. "The Focal Account: Indirect Lie Detection Need Not Access Unconscious, Implicit Knowledge." *Journal of Experimental Psychology: Applied* 21: 342–355.

Taylor, R., and R. E. Hick. 2007. "Believed Cues to Deception: Judgments in Self-Generated Trivial and Serious Situations." *Legal and Criminological Psychology* 12, no. 2: 321–331. https://doi.org/10.1348/135532506X116101.

Tversky, A., and D. Kahneman. 1974. "Judgment Under Uncertainty: Heuristics and Biases." *Science* 185, no. 4157: 1124–1131. https://doi.org/10.1126/science.185.4157.1124.

Ulatowski, J. 2018. "Is There a Commonsense Semantic Conception of Truth?" *Philosophia* 46, no. 2: 487–500. https://doi.org/10.1007/s11406-017-9941-x.

Verschuere, B., G. Bogaard, and E. Meijer. 2021. "Discriminating Deceptive From Truthful Statements Using the Verifiability Approach: A Meta-Analysis." *Applied Cognitive Psychology* 35, no. 2: 374–384. https://doi.org/10.1002/acp.3775.

Verschuere, B., C.-C. Lin, S. Huismann, et al. 2023. "The Use-the-Best Heuristic Facilitates Deception Detection." *Nature Human Behaviour* 7, no. 5: 5. https://doi.org/10.1038/s41562-023-01556-2.

Verschuere, B., M. Schutte, S. van Opzeeland, and I. Kool. 2021. "The Verifiability Approach to Deception Detection: A Preregistered Direct Replication of the Information Protocol Condition of Nahari, Vrij, and Fisher (2014b)." *Applied Cognitive Psychology* 35, no. 1: 308–316. https://doi.org/10.1002/acp.3769.

Vrij, A., K. Edward, and R. Bull. 2001. "Police officers' Ability to Detect Deceit: The Benefit of Indirect Deception Detection Measures." *Legal* *and Criminological Psychology* 6, no. 2: 185–196. https://doi.org/10.1348/135532501168271.

Vrij, A., and M. Hartwig. 2021. "Deception and Lie Detection in the Courtroom: The Effect of Defendants Wearing Medical Face Masks." *Journal of Applied Research in Memory and Cognition* 10, no. 3: 392–399. https://doi.org/10.1016/j.jarmac.2021.06.002.

Vrij, A., and F. W. Winkel. 1991. "Cultural Patterns in Dutch and Surinam Nonverbal Behavior: An Analysis of Simulated Police/Citizen Encounters." *Journal of Nonverbal Behavior* 15, no. 3: 169–184. https://doi.org/10.1007/bf01672219.

Vrij, A., and F. W. Winkel. 2011. "Statistical Evidence in Experimental Psychology: An Empirical Comparison Using 855 *t* Tests." *Perspectives on Psychological Science* 6, no. 3: 291–298. https://doi.org/10.1177/1745691611406923.

Vrij, A., C. A. Meissner, R. P. Fisher, S. M. Kassin, C. A. Morgan, and S. M. Kleinman. 2017. "Psychological Perspectives on Interrogation." *Perspectives on Psychological Science* 12, no. 6: 927–955. https://doi.org/10.1177/1745691617706515.

Way, K. 2021. "What's the Deal With Elizabeth Holmes 'Never Blinking'?" https://www.vice.com/en/article/xgxjdd/elizabeth-holmes-not-blinking.

Wetzels, R., D. Matzke, M. D. Lee, J. N. Rouder, G. J. Iverson, and E.-J. Wagenmakers. 2011. "Statistical Evidence in Experimental Psychology: An Empirical Comparison Using 855 *t* Tests." *Perspectives on Psychological Science* 6, no. 3: 291–298. https://doi.org/10.1177/1745691611406923.