CANADIAN PSYCHOLOGICAL ASSOCIATION
SOCIÉTÉ CANADIENNE DE PSYCHOLOGIE

Canadian Journal of Behavioural Science / Revue canadienne des sciences du comportement

ISSN: 0008-400X

2024, Vol. 56, No. 1, 31–40
https://doi.org/10.1037/cbs0000337

False Impressions? The Effect of Language Proficiency on Cues, Perceptions, and Lie Detection

Elizabeth Elliott[1] and Amy-May Leach[2]
[1] Department of Psychology, Iowa State University
[2] Faculty of Social Science and Humanities, University of Ontario Institute of Technology

We examined how observers' impressions of nonnative speakers and their cues to deception affected decision-making. Native and nonnative English speakers lied or told the truth about having committed a transgression, and then observers attempted to detect their lies. Observers were better able to discriminate between lie- and truth-telling native speakers than nonnative speakers. They also held more positive impressions of native speakers than nonnative speakers. Unlike observers, trained coders identified a multitude of differences in interviewees' presentation of cues to deception across proficiency groups. Overall, nonnative speakers appear to be at a significant disadvantage in lie-detection contexts.

Public Significance Statement
A decade of research on lie detection and language proficiency has revealed that a person's proficiency affects people's biases and their abilities to discriminate between lie-tellers and truth-tellers. We found that nonnative (vs. native) English speakers were viewed more negatively and exhibited different behaviours. These findings have significant implications for scholarship and practice given that this was the first study to examine the underpinnings of proficiency effects on deception detection.

Keywords: lie detection, decision-making, language proficiency, cues to deception, impressions

Witnesses' language proficiencies can affect lie detection accuracy and biases in certain contexts (e.g., Da Silva & Leach, 2013; Evans et al., 2013). However, researchers have yet to identify the underpinnings of these effects. Emerging research suggests that neither familiarity with nonnative speech (i.e., speaking the same language; Leach et al., 2017) nor task demands play a role (i.e., type of lie; Evans et al., 2017). Yet, anecdotal reports suggest that native speakers "look different" from nonnative speakers. We considered whether the overt behaviours of nonnative speakers' or observers' subjective impressions of the interviewees could account for proficiency effects.

Telling a lie is more cognitively taxing than telling the truth (Zuckerman et al., 1981). Lie-tellers must consciously suppress automatically generated accurate recall while fabricating accounts that are cohesive, plausible, and consistent (e.g., Spence et al., 2001; Walczyk et al., 2003). It is no surprise, then, that people are slower to formulate lies than they are to respond truthfully (Suchotzki & Gamer, 2018). Lie-tellers must also monitor their behaviours, and those of listeners, to ensure that their accounts are believed (DePaulo et al., 2003). In turn, the greater cognitive demands associated with lie telling (vs. truth telling) are revealed in behavioural cues to deception (DePaulo et al., 2003; Sporer & Schwandt, 2006; Zuckerman et al., 1981) and additional leads for interviewers (e.g., multiple cues to deception; Vrij, 2008).

Vrij, Fisher, et al. (2008) suggested that increasing cognitive load demands further, by having interviewees perform another task simultaneously (e.g., speaking a nonnative language), would highlight the differences between lie-tellers and truth-tellers. The imposition of two cognitively demanding tasks can cause an individual to alternate their attention between tasks and results in poor performance (Broadbent, 1957). For example, interviewees who were asked to speak in their nonnative language and lie performed these tasks slower than nonnative speakers who were instructed to tell the truth (Suchotzki & Gamer, 2018). It stands to reason, then, that the greater the cognitive demands placed on lie-tellers, the more difficulty they will have concealing their deception.

Ardila's (2003) review of language acquisition research revealed that nonnative speakers exhibited complex brain activation patterns during tasks involving working memory. Working memory efficiency is highest when individuals use their native, compared to nonnative, language (Baddeley, 2000). In fact, the task of speaking in a nonnative language is difficult even for professional interpreters (Rinne et al., 2000): Researchers found an increase in interpreters' brain activation when they translated accounts from their native to nonnative languages. Thus, lying and speaking a nonnative language may be doubly cognitively taxing, leading to verbal or nonverbal leaks and easier detectability.

This article was published Online First June 9, 2022.
Elizabeth Elliott https://orcid.org/0000-0002-9655-1151
Amy-May Leach https://orcid.org/0000-0002-9464-1284
The authors would like to acknowledge Erin Billinger for her contributions to the study. The authors have no known conflict of interest to disclose.
The data that support the findings of this study are available from the corresponding author upon reasonable request.
Correspondence concerning this article should be addressed to Elizabeth Elliott, Department of Psychology, Iowa State University, 1347 Stange Road, Lagomarcino Hall, Ames, IA 50011, United States. Email: elliotte@iastate.edu

Language Proficiency and Lie Detection

Accuracy

Research on the role of language proficiency in lie detection has been relatively sparse and has yielded mixed findings (i.e., every possible result has been found). In the first study of its kind, no differences were found in observers' abilities to detect native and nonnative speakers' deception (e.g., opinions about abortion; Cheng & Broadhurst, 2005). This pattern of null results was subsequently replicated in different contexts (Castillo et al., 2014; Evans & Michael, 2014).

Yet, nonnative speakers' levels of language proficiency appear to be a significant factor. Observers were better at detecting deception in high- and low-proficiency nonnative English speakers than native English speakers, for example (Evans et al., 2013). Indeed, observers performed best when discriminating between lie- and truth-tellers in the lowest proficiency group (vs. all other levels of proficiency; Evans et al., 2017). These effects held true even when written messages of nonnative speakers were judged (Volz et al., 2020). On the whole, these findings suggest that proficiency level might have an impact on accuracy, with observers being better able to detect nonnative speakers' lies.

When nonnative speakers have been recruited from language centres, proficiency effects have been reported, as well. However, the direction of effects has been reversed. In three studies, featuring the same interviewees, police officers (Leach & Da Silva, 2013) and untrained observers (Da Silva & Leach, 2013; Leach et al., 2017) were better able to detect deception in native than nonnative speakers. Similarly, observers were better able to differentiate between falsified and truthful identities provided in a native (vs. nonnative) language (Akehurst et al., 2018).

Observers' performance appears to be linked to interviewees' levels of proficiency. Observers' abilities to discriminate between lie- and truth-telling mock witnesses have been poorest when judging beginner English speakers than any of the other proficiency groups (i.e., native, advanced, or intermediate English speakers; Elliott & Leach, 2016).

Bias

Unlike accuracy, there has been remarkable consistency across studies in terms of observers' decision biases. Typically, a truth bias has been elicited by native speakers, but not nonnative speakers (e.g., Elliott & Leach, 2016; Evans et al., 2017; Evans & Michael, 2014). In fact, a truth bias towards nonnative speakers has only been found once, but the nonnative speakers were speaking in their native language (i.e., code switched; Bond & Atoum, 2000). Most often, however, nonnative speakers have either elicited no bias or a lie bias (e.g., Castillo et al., 2014; Da Silva & Leach, 2013). Thus, findings indicate that nonnative speakers are perceived less positively than native speakers.

Cues to Deception

Speakers' cues could account for proficiency-related differences in lie detection. Both untrained observers and police officers hold similar stereotypes of deceivers, regardless of their language proficiencies (Leach et al., 2019). Similarly, in an explicit lie detection test, observers stated that they used the same cues to detect deception in native and nonnative speakers (Cheng & Broadhurst, 2005). There were differences in the cues that lie- and truth-tellers reported displaying, however. Lie-tellers noted experiencing more difficulty controlling verbal and nonverbal behaviours (e.g., leg and foot movements, vocal pitch) than truth-tellers. Interestingly, these self-reported difficulties were more pronounced among nonnative lie-tellers than native-speaking lie-tellers.

When observers have employed the psychologically based credibility assessment tool (PBCAT), more cues to deception were noted in nonnative than native speakers (e.g., Evans et al., 2013). However, these minimally trained observers primarily used the PBCAT as a decision-making aid and did not objectively code cues. Moreover, the PBCAT items were not completely comprehensive because the tool only consisted of 11 cues to deception. Thus, it remains unclear the extent to which a full range of nonverbal and verbal behaviours are actually exhibited by lie- and truth-telling nonnative speakers.

Recent research suggests that diagnostic deception cues can be faint and inconsistent (for a full discussion of the lens model, see Hartwig & Bond, 2011). Special training or equipment might be needed to truly detect whether cues to deception are present. There has only been one study in which native and nonnative speakers' lie- and truth-telling behaviours were examined by trained coders. Elliott and Leach (2016) found that nonnative speakers exhibited more speech disturbances (e.g., repetitions) than native speakers. However, unlike with the PBCAT, lying in a nonnative language did not affect cues related to emotion (e.g., nervousness) or cognitive load (e.g., thinking hard). It is possible that the playback rate (i.e., normal speed rather than frame by frame) and coding method (i.e., use of hand calculations rather than specialized coding software) might have contributed to Type II error. Thus, more precise measures should be utilized to obtain a comprehensive account of proficiency effects.

Impressions

Observers' impressions of nonnative speakers might also affect their decision-making. Native language accents are perceived more favourably than nonnative accents (Gill, 1994). For example, stronger nonnative accents have been associated with perceptions of lower intellectual abilities and comprehension. Observers also pay less attention to the details provided by nonnative speakers because they expect them to communicate poorly (Lev-Ari & Keysar, 2012). In turn, accents can impact believability. In one study, ratings of the statements' truthfulness decreased as readers' accents became stronger (Lev-Ari & Keysar, 2010). Observers seemed to be basing their judgements of deception on their impressions of the readers.

Negative impressions of nonnative speakers may also arise in the absence of speech. Nonnative speakers tend to hesitate by pausing silently as they think (Temple, 1992), whereas native speakers use fillers (e.g., "umm"; Maclay & Osgood, 1959). This difference is problematic because hesitations (e.g., pausing) have been linked to low intelligibility (Frayer & Krasinski, 1995), incomprehensibility (Blau, 1991), and attributions of guilt (Hosman & Wright, 1987). Nonnative speakers might, therefore, unknowingly exhibit cues that might lead observers to perceive them negatively.

The implications of observers' impressions of nonnative speakers during lie-detection decision-making have only recently begun to be explored. Castillo et al. (2014) found that native English observers deemed native Spanish speakers (i.e., nonnative English speakers) less trustworthy than native English speakers. However, the results of this study are limited in that observers made decisions about interviewees who spoke different languages (i.e., English and Spanish) rather than those with varying levels of proficiency within a single language (e.g., English). Thus, the impact of language proficiency on impressions requires additional study.

The Present Study

We examined whether interviewees' cues to deception and observers' impressions could account for effects of language proficiency on lie detection. Here, we compared native English speakers and beginner nonnative English speakers. We chose a low-proficiency nonnative sample because it is a rarely examined sample, yet low-proficiency nonnative speakers are viewed most negatively (compared to native speakers; i.e., Elliott & Leach, 2016), and may, therefore, be placed at a greater disadvantage in the justice system. Based on these findings and research on nonnative speakers' working memory (Ardila, 2003; Baddeley, 2000), we expected to replicate previous findings and hypothesized that observers would be better able to detect deception in native than nonnative speakers. We also expected observers to have a greater truth bias towards native speakers than nonnative speakers (e.g., Evans et al., 2017).

We examined the underpinnings of these effects in two ways. First, interviewees' verbal and nonverbal behaviours were analysed by observers and trained coders. That is, observers reported the cues that they used to make their lie detection decisions, and nonverbal cues were independently verified with frame-by-frame analyses of interviewees' accounts using coding software (i.e., Datavyu; Datavyu Team, 2014). We hypothesized that untrained observers would be relatively insensitive to differences between native and nonnative speakers (e.g., Leach et al., 2019). However, trained coders were expected to find that nonnative speakers exhibited more cues to deception than native speakers (e.g., Elliott & Leach, 2016). Second, we examined observers' impressions of each interviewee. In keeping with hypotheses regarding bias, we expected observers to perceive nonnative speakers more negatively than native speakers, particularly when they were lying.

Method

Participants

An a priori power analysis (i.e., fixed-effects, omnibus, one-way) revealed that 122 participants were needed to achieve a power of 80% with a medium effect size. A total of 123 undergraduates at Ontario Tech University (females = 77, males = 46; M_{age} = 20.43 years, SD_{age} = 3.50 years) participated in the study in exchange for extra credit. Observers self-identified as South Asian (29.3%), Caucasian (23.6%), Black (17.1%), Arab (14.6%), Other (5.7%), South East Asian (4.1%), Chinese (2.4%), Aboriginal (.8%), Hispanic (.8%), Korean (.8%), and Latin American (.8%). Most of them stated that their native language was English (72.4%) and rated themselves as "very fluent" in English (81.3%).[1]

Materials

Videos

We used hidden camera footage, in which participants' upper bodies and facial features were visible, that was obtained from Da Silva and Leach (2013). They used a modified version of Russano et al.'s (2015) cheating paradigm.

Upon arrival at the laboratory, each interviewee was seated next to a confederate who was posing as a student. A female experimenter instructed both "participants" to complete a set of logic problems individually. Before leaving the room for 15 min, she emphasized that they were not to aid each other with responses. The confederate was randomly assigned to ask half of the interviewees for help with one of the questions while the experimenter was out of the room (i.e., the experimenter was blind to condition). The experimenter looked worried upon returning to collect the logic problems before excusing herself again. While she was gone, the confederate alerted the interviewee to the experimenter's worry and asked what the interviewee thought might be wrong. In the cheating condition, she also asked the interviewee not to divulge that they had shared their answers.

When the experimenter returned, she announced that the pair had the same wrong answer on the test. She sent the confederate to the waiting room and told the interviewee that he or she would be questioned first. The experimenter proceeded to interrogate the interviewee by asking a series of open- and close-ended questions: "What do you think the problem is?" "I left the room for fifteen minutes; describe everything that happened from the minute I left until I returned" "Can you be more specific? I really need to know what happened?" "Did you ask the other student for help?" "Did she ask you for help?" "Did you share your answers?" "While I was gone, did you cheat on the test?" "What do you think we should do about this?" After the final response, the participant received the manipulation questionnaire, and then they were debriefed.

Overall, we used clips of 10 truth-tellers and five lie-tellers per proficiency group (M_{length} = 139 s, SD = 55.04 s, 50% female). Lie-tellers typically provided denials in response to accusations of cheating.

Interviewees' language proficiencies were determined using their responses to a language history questionnaire[2] (LHQ; Li et al., 2006). For example, they were asked to rate their English fluency using a Likert scale (i.e., 1 = *not fluent at all* to 5 = *extremely fluent*). Only those interviewees who reported being "extremely fluent" were placed in the native English speaker condition. Native English speakers (n = 15) were recruited from Ontario Tech University, whereas nonnative speakers (n = 15) were recruited from the campus English learning centre (see Leach et al., 2017, for a full description). Their English-speaking, listening, reading, and writing skills had been evaluated by the centre using a standardized measure of proficiency (i.e., Canadian Language Benchmarks, 2012). Based on this measure, nonnative speakers had been classified as beginner English speakers (i.e., scored 4 or less on a 1–12 scale).

[1] Observers' language proficiency does not impact discrimination (Leach et al., 2017).

[2] LHQs have been used in similar studies to categorize interviewees (e.g., Evans et al., 2013).

Lie Detection Questionnaire

Observers indicated whether each interviewee was lying or telling the truth, and how confident they were in their decision (from 0% = *not at all confident* to 100% = *extremely confident*).[3]

Cues Questionnaire

Observers were asked about the cues they used to make lie detection decisions (i.e., "What cues did you use to make your decision?"). Then, they were instructed to indicate the extent to which they believed each of the cues listed had increased or decreased during deception using a Likert scale (i.e., −3 = *decreasing during deception*, 0 = *no change/not used*, 3 = *increasing during deception*). Observers were instructed to provide ratings for all cues listed. To avoid influencing observers' decisions, we provided a wide array of verbal, vocal, and nonverbal cues from DePaulo et al. (2003) meta-analysis of cues to deception—even those that were known to be nondiagnostic: blinking, coherence, inconsistency, cooperativeness, spontaneous corrections, covering mouth/eyes, amount of detail, grammatical errors, eye contact, facial expressiveness, fidgeting, generalizations, illustrators, hesitations, duration of answers, length of pauses, leg/foot movements, lack of memory, negative statements, nervousness, pauses, vocal pitch, pupil dilation, rate of speech, repetition, self-manipulation, shifts in posture, smiling, stuttering, vocal tension, unusual detail, unfriendly facial expression, and vagueness.

Cue Coding

Trained coders used transcripts and a video coding software (i.e., Datavyu; Datavyu Team, 2014) to analyse all cues. Datavyu was used to analyse nonverbal cues (i.e., chin raises, fidgeting, lip presses, pauses, response length, smiling, and unfriendly facial expression) frame by frame. Coding the frame onset (e.g., start of lip movement to form smile) and offset (e.g., return to neutral position) of cues allowed us to generate information about the frequencies and durations of cues (all cues except chin raises and lip presses). Next, coders noted the frequency of the following verbal cues using interview transcripts: admitted lack of memory, details, grammatical errors, negative statements, speech disturbances, spontaneous corrections, uncertainty, and word or phrase repetition. Finally, coders watched each interview and made global judgements of cues that could not be measured in terms of frequencies or duration (i.e., coherent, cooperative, discrepant, nervous, plausible, vocal pitch, and vocal tension) using a Likert scale (1 = *not at all* to 5 = *very/throughout*).

One coder rated all cues and four coders rated 25% each of all material, any disagreements were resolved through conversation. Using the two-way random-effects model (see Koo & Li, 2016), interrater reliability was adequate (i.e., interclass correlation coefficients [ICCs] > .60 for nervousness, vocal tension, and pitch) to high (i.e., all other ICCs > .75). Coders rated one cue at a time for all interviews before coding the next cue. The coding scheme was based on significant indicators of deception (DePaulo et al., 2003; Sporer & Schwandt, 2006; Sporer & Schwandt, 2007). We also selected cognitive load cues (e.g., Vrij, Mann, et al., 2008) and cues that could be affected by proficiency (e.g., Sert & Jacknick, 2015; Tavakoli & Foster, 2011).

Impressions Questionnaire

Observers provided their impressions of each interviewee. Specifically, they indicated the extent to which they agreed (i.e., 1 = *strongly agree* to 5 = *strongly disagree*) that each interviewee exhibited a list of following characteristics: articulate, believable, capable, confident, intelligent, intentionally lying, likable, relaxed, sincere, thinking hard, and trustworthy. These impressions were based on previous lie detection, impression formation, and cross-cultural studies (e.g., Leary & Kowalski, 1990; Vrij & Winkel, 1991, 1994).

Procedure

After we gained approval from our institutional research ethics board, we recruited undergraduate students using our university's online participant pool (i.e., Sona Systems; https://www.sona-systems.com/). Participants (henceforth referred to as "observers") were seated at a computer in a quiet room. They were randomly assigned to view 15 lie-tellers and truth-tellers who were either native or nonnative English speakers. Video order was counterbalanced (i.e., half of the observers viewed videos in a reverse order—from the last video to the first) to account for order effects. Following each video, observers rendered their lie detection decisions. Half of the observers rated their impressions of the interviewee after each video and the other half identified the cues that they used to make their lie detection decisions overall (i.e., after viewing all the videos). Observers completed a demographics questionnaire at the end of the session.

Results

Preliminary analyses indicated nonsignificant effects, thus we collapsed across the following variables: gender, age, race, English fluency, years speaking English, experience interacting with nonnative speakers, language spoken at home, and native language.

Untrained Observers

Signal Detection Theory

We used signal detection theory (SDT; Green & Swets, 1966) and Stanislaw and Todorov's (1999) formulas to conduct analyses on observers' abilities to discriminate between lie- and truth-tellers (i.e., $d' = Z_{\text{hit}} - Z_{\text{false alarm}}$) and their decision bias (i.e., $c = -[(Z_{\text{hit}} + Z_{\text{false alarm}})]/2$). A decision was categorized as a "hit" if the observer correctly identified a lie-teller, whereas classifying a truth-teller as a lie-teller was considered a "false alarm" (FA). The SDT formulas accounted for the uneven numbers of lie- and truth-tellers (i.e., 5 vs. 10 interviewees) in the calculation of false alarms (i.e., FA = number of interviewees identified as lie-tellers/10) and hits (H = number of interviewees identified as lie-tellers/5). We also applied a standard correction to all extreme values (i.e., 0 s to .5/signal or noise, and 1 s to [signal or noise − .5]/signal or noise). Thus, hit rates of 0 were changed to .10, hit rates of 1 were changed to .90, false alarms of 0 were changed to .05, and false alarms of 1 were changed to .95.

Discrimination. We conducted a one-way analysis of variance (ANOVA) on observers' abilities to discriminate between lie- and

[3] Confidence results are available from the corresponding author.

truth-tellers. As hypothesized, observers were better able to discriminate between lie- and truth-telling native speakers ($M = .55$, $SD = .74$) than nonnative English speakers ($M = -.14$, $SD = .88$), $F(1, 122) = 22.03$, $p < .001$, $\eta_p^2 = .15$, 95% CI [.05, .27]. We then used one-sample t tests to compare observers' discrimination scores to "0" (i.e., no discrimination). Observers were able to discriminate between lie- and truth-tellers who were native speakers, $t(60) = 5.79$, $p < .001$, $d = .74$, 95% CI [.46, 1.02], but not nonnative speakers, $t(61) = 1.23$, $p = .225$, $d = .16$, 95% CI [.09, .41].

Bias. A one-way ANOVA was conducted on observers' response bias. There was no effect of proficiency, $F(1, 122) = 2.48$, $p = .118$, $\eta_p^2 = .10$, 95% CI [.00, .35]. We used one-sample t tests to compare observers' response bias to "0" (i.e., no bias). Observers exhibited a truth bias towards native speakers ($M = .17$, $SD = .48$), $t(60) = 2.70$, $p = .009$, $d = .34$, 95% CI [.09, .60], and nonnative speakers ($M = .29$, $SD = .43$), $t(61) = 5.41$, $p < .001$, $d = .69$, 95% CI [.41, .96].

Cues

We conducted a multivariate analysis of variance (MANOVA) on observers' reported beliefs about the increase or decrease of cues during deception. As expected, there was no effect of proficiency, $F(1, 58) = 1.23$, $p = .302$, $\eta_p^2 = .02$, 95% CI [.00, .14], on the combined dependent variables. Effects of veracity were not examined because observers reported the deception cues that they used to make their decisions after they watched all of the videos (i.e., for both lie and truth).

Impressions

A mixed-factors MANOVA was conducted on observers' impressions (see Table 1, for Ms and SDs; $\alpha_{altered} = .005$). We found a main effect of proficiency on the combined dependent variables, $F(1, 58) = 2.79$, $p = .007$, $\eta_p^2 = .05$, 95% CI [.00, .18]. Univariate analyses revealed that only interviewees' impressions of being *articulate* were significant; however, this effect was qualified by a higher order interaction. There was also a main effect of veracity, $F(1, 58) = 10.97$, $p < .001$, $\eta_p^2 = .16$, 95% CI [.03, .32], on the combined dependent variables. Compared to lie-tellers, truth-tellers were rated as more *believable*, $F(1, 58) = 19.99$, $p < .001$, $\eta_p^2 = .26$, 95% CI [.08, .42], *capable*, $F(1, 58) = 58.37$, $p < .001$, $\eta_p^2 = .50$, 95% CI [.31, .63], *intelligent*, $F(1, 58) = 11.45$, $p = .001$, $\eta_p^2 = .16$, 95% CI [.03, .33], *likable*, $F(1, 58) = 8.98$, $p = .004$, $\eta_p^2 = .13$, 95% CI [.01, .30], *relaxed*, $F(1, 58) = 22.57$, $p < .001$, $\eta_p^2 = .28$, 95% CI [.10, .44], and *trustworthy*, $F(1, 58) = 9.64$, $p = .003$, $\eta_p^2 = .14$, 95% CI [.02, .31]. Lie-tellers were viewed as *thinking hard(er)* than truth-tellers, $F(1, 58) = 22.20$, $p < .001$, $\eta_p^2 = .28$, 95% CI [.10, .44]. We found effects of veracity on impressions of *intent to lie* and *sincerity* as well, but these were qualified by interactions.

Indeed, there were several interactions between the variables, $F(1, 58) = 4.96$, $p < .001$, $\eta_p^2 = .08$, 95% CI [.00, .23], specifically in terms of the following impressions: *articulate*, $F(1, 58) = 10.44$, $p = .002$, $\eta_p^2 = .15$, 95% CI [.02, .32], *intention to lie*, $F(1, 58) = 12.42$, $p = .001$, $\eta_p^2 = .18$, 95% CI [.03, .34], and *sincerity*, $F(1, 58) = 14.60$, $p < .001$, $\eta_p^2 = .20$, 95% CI [.05, .37]. Post hoc independent-samples t tests revealed that, when lying, native speakers were rated as more *articulate* than nonnative speakers, $t(59) = 4.28$, $p < .001$, $d = 1.11$, 95% CI [.56, 1.64]; native and nonnative speakers were perceived as equally *articulate* when they were telling the truth, $t(59) = 1.00$, $p = .320$, $d = .26$, 95% CI [−.25, .77]. In terms of interviewees' *intent to lie*, there was no difference between native and nonnative lie-tellers, $t(59) = 1.80$, $p = .078$, $d = .46$, 95% CI [−.05, .98]; amongst truth-tellers, observers perceived nonnative speakers as having a greater *intent to lie* than native speakers, $t(59) = 2.63$, $p = .011$, $d = .68$, 95% CI [.16, 1.20]. Finally, truth-tellers were perceived to be equally *sincere* regardless of proficiency, $t(59) = .48$, $p = .633$, $d = .12$, 95% CI [−.38, .63], whereas lie-tellers who were native speakers were regarded as more *sincere* than nonnative speakers, $t(59) = 2.62$, $p = .012$, $d = .68$, 95% CI [.15, 1.19].

Table 1
Observers' Impressions of Interviewees by Veracity and Proficiency

	Truth-tellers (*M*/*SD*)		Lie-tellers (*M*/*SD*)	
Impression	Native	Nonnative	Native	Nonnative
Articulate	2.63/.48	2.75/.37	2.80/.39	3.31/.52
Believable	1.94/.47	2.24/.56	2.13/.49	2.56/.71
Capable	2.62/.42	2.79/.59	2.95/.50	3.35/.59
Confident	2.39/.47	2.40/.48	2.97/.54	3.30/.54
Intelligent	2.51/.39	2.70/.42	2.61/.44	2.93/.54
Intentionally lying	3.29/.48	2.95/.52	2.89/.44	3.09/.43
Likable	2.43/.38	2.43/.44	2.69/.46	2.54/.52
Relaxed	2.68/.59	2.69/.55	2.92/.68	3.15/.66
Sincere	2.53/.52	2.59/.56	2.96/.49	2.54/.72
Thinking hard	3.33/.55	2.99/.64	2.95/.57	2.75/.67
Trustworthy	2.64/.34	2.76/.48	2.94/.46	2.83/.55

Note. M = mean; SD = standard deviation. The impression scale ranged from 1 (*strongly agree*) to 5 (*strongly disagree*).

Trained Coders

To account for differences in interview length across interviewees, we recoded all cues (except global judgements) by dividing the raw score (on each cue) by the total duration of the video (in seconds). This correction was completed to ensure that the prevalence of cues was disassociated from the length of videos. See Table 2, for means and standard deviations for each cue, and Table 3 for the results of the MANOVAs and subsequent univariate analyses. We conducted Veracity × Proficiency MANOVAs on these recoded assessments of interviewees' deception cues. We employed a Bonferroni correction—$\alpha_{altered}$—to examine the univariate analyses.

Frequency of Cues

We found a main effect of proficiency, $F(1, 29) = 2.51$, $p = .047$, $\eta_p^2 = .08$, 95% CI [.00, .29], on the combined dependent variables. An examination of the univariate effects ($\alpha_{altered} = .006$) revealed that nonnative speakers *paused* more often than native speakers, $F(1, 29) = 9.72$, $p = .004$, $\eta_p^2 = .25$, 95% CI [.03, .47], whereas native speakers provided more *detail* than nonnative speakers, $F(1, 29) = 33.65$, $p < .001$, $\eta_p^2 = .54$, 95% CI [.26, .69].

Duration of Cues

We found an effect of proficiency, $F(1, 29) = 4.09$, $p = .006$, $\eta_p^2 = .12$, 95% CI [.00, .34]. The univariate analysis ($\alpha_{altered} = .007$) revealed that nonnative speakers *paused* longer, $F(1, 29) = 11.23$, $p = .002$, $\eta_p^2 = .28$, 95% CI [.04, .49], and *smiled* longer, $F(1, 29) =$

Table 2
Trained Coders' Analysis of Cues Exhibited by Interviewees

Cue[a]	Coding method	Truth-tellers (*M*/*SD*)		Lie-tellers (*M*/*SD*)	
		Native	Nonnative	Native	Nonnative
Arm movements	Frequency	.14/.08	.11/.06	.08/.04	.08/.04
	Duration	.30/.15	.27/.18	.30/.21	.23/.15
Chin raises	Frequency	.00/.00	.00/.00	.00/.00	.00/.00
Eye contact	Frequency	.22/.07	.20/.09	.26/.07	.19/.04
	Duration	.65/.14	.68/.10	.53/.14	.71/.11
Fidgeting	Frequency	.09/.06	.06/.04	.05/.03	.03/.02
	Duration	.34/.30	.34/.37	.33/.31	.51/.37
Lip presses	Frequency	.04/.03	.07/.05	.07/.05	.05/.04
Pauses	Frequency	.09/.04	.13/.04	.08/.02	.12/.01
	Duration	.06/.02	.10/.06	.06/.02	.13/.06
Response	Frequency	.11/.04	.14/.04	.11/.05	.12/.01
	Duration	.59/.12	.48/.17	.62/.04	.44/.04
Smiling	Frequency	.08/.05	.12/.03	.07/.06	.12/.02
	Duration	.24/.26	.39/.14	.09/.09	.49/.28
Unfriendly facial expression	Frequency	.07/.05	.09/.06	.09/.05	.05/.03
	Duration	.10/.14	.13/.10	.12/.16	.07/.04
Admitted lack of memory	Transcript	.01/.01	.00/.00	.01/.01	.01/.01
Details	Transcript	.12/.05	.06/.02	.19/.09	.04/.01
Grammatical errors	Transcript	.04/.03	.02/.01	.01/.01	.01/.01
Negative statements	Transcript	.01/.01	.01/.01	.00/.00	.02/.01
Speech disturbances	Transcript	.03/.02	.05/.03	.04/.02	.04/.03
Spontaneous corrections	Transcript	.03/.03	.02/.02	.04/.02	.02/.01
Uncertainty	Transcript	.02/.01	.02/.02	.02/.01	.03/.03
Word or phrase repetition	Transcript	.02/.01	.03/.02	.02/.01	.02/.02
Coherence	Global judgements	4.80/0.35	3.50/0.53	4.40/0.89	2.60/1.34
Cooperation	Global judgements	3.50/0.75	3.15/0.78	4.20/0.45	2.20/1.09
Discrepancy	Global judgements	3.40/0.94	2.60/1.24	4.20/0.45	2.20/1.10
Nervousness	Global judgements	2.80/0.48	3.00/0.62	2.70/0.67	2.40/0.96
Plausibility	Global judgements	5.00/0.00	4.15/0.82	4.80/0.45	3.30/1.30
Vocal pitch	Global judgements	2.95/0.16	3.00/0.00	3.00/0.00	3.00/0.00
Vocal tension	Global judgements	3.50/0.58	2.95/1.21	3.20/0.84	3.60/1.19

Note. *M* = mean; *SD* = standard deviation. Global judgements of interviewee accounts were made using a Likert scale (1 = *not at all* to 5 = *very/throughout*).
[a] All cues (except global judgements) presented are calculated ratings (i.e., raw score dividing by length of video).

11.64, $p = .002$, $\eta_p^2 = .29$, 95% CI [.05, .50], but gave shorter *responses*, $F(1, 29) = 9.24$, $p = .005$, $\eta_p^2 = .24$ 95% CI [.03, .46] than native speakers.

Global Judgements

We found a main effect of proficiency, $F(1, 29) = 6.87$, $p < .001$, $\eta_p^2 = .19$, 95% CI [.01, .41]. A closer examination ($\alpha_{altered} = .007$) revealed that the accounts of native speakers were more *discrepant*, $F(1, 29) = 12.40$, $p = .002$, $\eta_p^2 = .30$, 95% CI [.05, .51], *coherent*, $F(1, 29) = 29.75$, $p < .001$, $\eta_p^2 = .51$, 95% CI [.23, .66], and *plausible*, $F(1, 29) = 17.56$, $p < .001$, $\eta_p^2 = .38$, 95% CI [.11, .57], than those of nonnative speakers (results presented in Table 3).

Discussion

We replicated previously reported effects of proficiency on lie detection. As expected, observers were better able to discriminate between lie- and truth-telling native English speakers than nonnative English speakers (e.g., Da Silva & Leach, 2013). They were specifically better at comparing between lie- and truth-tellers who were native English speakers but could not discriminate between lie- and truth-tellers who were nonnative English speakers. These findings add to the literature supporting the notion that people are more easily able to determine whether a native English speaker is lying or telling the truth compared to when they observe a nonnative English speaker (e.g., Akehurst et al., 2018; Elliott & Leach, 2016).

Response biases were evident in response to both native and nonnative English speakers. Observers tended to believe that native speakers were telling the truth. This truth bias has been well established in the literature (e.g., Bond & DePaulo, 2006). The truth bias towards nonnative English speakers is more difficult to explain. Typically, there has been a lie bias or no bias towards interviewees with this level of proficiency (e.g., Da Silva & Leach, 2013; Leach & Da Silva, 2013). However, this result has been found once before in a study where nonnative English speakers were permitted to code-switch (i.e., speaking their native language; Bond & Atoum, 2000). The finding may be unusual, but it does replicate previous work.

Our analyses of cues to deception yielded interesting results. Observers reported that native and nonnative speakers exhibited similar cues to deception. Given that untrained observers' and police officers' stereotypes of deceivers are unaffected by language proficiency (e.g., Leach et al., 2019), this result was not entirely unexpected. Stereotypes of deception may be so salient that they are applied overly broadly and are impervious to proficiency. Conversely, cues to deception themselves may be too difficult to discern without specialized training (e.g., Hartwig & Bond, 2011).

Indeed, group differences emerged when we employed trained coders. Specifically, nine cues differentiated between nonnative

Table 3

Results of the 2 × 2 Multivariate Analyses of Variances (MANOVAs) on Trained Coders' Cue Analyses

Measure	Veracity			Proficiency			Veracity × Proficiency		
	$F(1, 29)$	p	η_p^2	$F(1, 29)$	p	η_p^2	$F(1, 29)$	p	η_p^2
Frequency of cues[a]	1.03	.451	.30	**2.51**	**.047**	**.51**	0.60	.769	.20
Arm gestures	2.89	.101	.10	0.52	.476	.02	0.34	.566	.01
Eye contact	0.19	.668	.01	2.38	.135	.08	0.94	.340	.04
Fidgeting	2.81	.106	.10	1.55	.224	.06	0.12	.736	.00
Lip presses	0.20	.655	.01	0.02	.885	.00	2.25	.146	.08
Pauses	0.47	.500	.02	**9.72**	**.004**	**.27**	0.03	.873	.00
Responses	0.62	.440	.02	3.71	.065	.13	0.40	.532	.02
Smiling	0.17	.687	.01	4.44	.045	.15	0.00	.964	.00
Unfriendly facial expression	0.16	.693	.01	0.16	.689	.01	1.47	.236	.05
Duration of cues[b]	0.31	.941	.09	**4.09**	**.006**	**.59**	1.09	.405	.28
Arm gestures	0.06	.817	.00	0.59	.450	.02	0.09	.763	.00
Eye contact	1.05	.315	.04	4.77	.038	.16	2.66	.115	.09
Fidgeting	0.38	.546	.01	0.42	.524	.02	0.48	.496	.02
Pause duration	0.72	.403	.03	**11.23**	**.002**	**.30**	1.62	.214	.06
Response length	0.02	.889	.00	**9.24**	**.005**	**.26**	0.73	.402	.03
Smile duration	0.11	.742	.00	**11.64**	**.002**	**.31**	2.30	.141	.08
Unfriendly facial expression	0.19	.670	.01	0.09	.768	.00	0.71	.409	.03
Transcript analysis[a]	0.82	.595	.26	**6.41**	**<.001**	**.73**	1.20	.352	.34
Admitted lack of memory	1.16	.292	.04	0.39	.536	.02	0.51	.480	.02
Details	2.48	.128	.09	**33.65**	**<.001**	**.56**	5.50	.027	.18
Grammatical errors	4.83	.037	.16	1.21	.282	.04	2.03	.166	.07
Negative statements	0.12	.734	.01	5.88	.023	.18	3.57	.070	.12
Speech disturbances	0.08	.778	.00	1.60	.219	.06	1.33	.259	.05
Spontaneous corrections	0.06	.810	.00	2.06	.163	.07	0.60	.445	.02
Uncertainty	0.10	.752	.00	0.52	.477	.02	0.60	.447	.02
Word or phrase repetition	0.17	.686	.01	1.72	.201	.06	0.06	.815	.00
Global judgements[b]	1.35	.278	.32	**6.87**	**<.001**	**.71**	1.00	.463	.26
Coherent	5.23	.031	.17	**29.75**	**<.001**	**.53**	0.77	.387	.03
Cooperative	0.06	.804	.00	2.52	.125	.09	0.17	.680	.01
Discrepancy	0.25	.619	.01	**12.40**	**.002**	**.32**	2.28	.143	.08
Nervousness	1.91	.178	.07	0.04	.845	.00	0.98	.332	.04
Vocal pitch	0.48	.494	.02	0.48	.494	.02	0.48	.494	.02
Plausibility	3.51	.072	.12	**17.56**	**<.001**	**.40**	1.34	.257	.05
Vocal tension	0.22	.647	.01	0.04	.844	.00	1.58	.220	.06

Note. The table shows results for main effects and interactions on the combined dependent variables (i.e., coding method), subsequent univariate analyses are also presented. Significant results are in boldface (according to corresponding adjusted α), cues are in italics. Interclass correlation coefficients (ICCs) for each cue were as follows: arm gestures (ICC = .77), eye contact (ICC = .86), fidgeting (ICC = .82), lip presses (ICC = .93), pauses (ICC = .91), responses (ICC = .94), smiling (ICC = .97), admitted lack of memory (ICC = .96), details (ICC = .87), grammatical errors (ICC = .83), negative statements (ICC = .79), speech disturbances (ICC = .78), spontaneous corrections (ICC = .87), uncertainty (ICC = .94), word or phrase repetition (ICC = .98), coherent (ICC = .89), cooperative (ICC = .82), discrepant (ICC = .84), nervousness (ICC = .67), vocal pitch (ICC = .61), plausibility (ICC = .80), and vocal tension (ICC = .66).
[a] Bonferroni-adjusted α .006. [b] Bonferroni-adjusted α .007.

speakers and native speakers. Contrary to our expectations, however, these differences were not exacerbated by deception. In previous work with the PBCAT, lower proficiency nonnative speakers exhibited more cues (e.g., details, spontaneous corrections, coherence, negative statements) than native speakers (Evans et al., 2013). Regardless, the differences between native and nonnative English speakers' cues might have influenced decision-making. The experience of interviewees with the nonnative proficiencies should differ most from native speakers in terms of cognitive demands and related behavioural effects (Evans et al., 2017; Gregersen, 2005). That is precisely what occurred. Interestingly, three of the cues that were more likely to be exhibited by nonnative speakers—longer pauses, shorter responses, and greater discrepancies—are associated with the prototype of a deceiver (Global Deception Research Team, 2006). The inappropriate application of a worldwide stereotype might explain why observers were less able to detect the deception of this group.

Observers' impressions also depended on interviewees' proficiencies and veracity. As predicted, lie-tellers elicited more negative impressions than truth-tellers. Researchers have reported similar unfavourable judgements (e.g., unlikable, untrustworthy) about lie-tellers in previous research (e.g., van 't Veer et al., 2015). We also found unfavourable impressions of nonnative compared to native English speakers on several items (i.e., articulate, sincere, intent to lie). One explanation might be that accents elicit more negative impressions (Gill, 1994; Lev-Ari & Keysar, 2010). Furthermore, we expected that combining nonnative speech with lie telling would increase negativity. We found three meaningful interactions between proficiency and veracity. Observers perceived native speakers as being equally articulate as nonnative speakers when they were telling the truth, but more articulate when native speakers lied. Although proficiency did not highlight differences in perceptions of sincerity among truth-tellers, native English speakers were seen as more sincere in their accounts than nonnative English

speakers when they lied. Evidently, impressions of truth-tellers were similar across proficiencies except in relation to their intent to lie. That is, truth-telling nonnative English speakers were more likely to be perceived as intentionally lying than native English speakers, but there were no differences in perceived intent to lie in either proficiency groups of lie-tellers. On the whole, our findings suggest that different cues and impressions might underlie proficiency differences in lie detection and that speaking a nonnative language may account for the observed differences in accuracy.

Limitations and Future Research

Our results are limited by the paradigm that we employed to elicit deceptive responses. Recall that we used a modified Russano et al. (2015) cheating paradigm. We chose this paradigm because it provides psychological realism, although it is lacking in cognitive demands and often results in short responses (e.g., denials). Other paradigms, such as the mission paradigm (see Vrij et al., 2010, for a description), might have allowed interviewees to provide lengthier, more creative responses and, thus, afforded them more chances to display cues and influence observers' impressions (Vrij, 2008). We do not believe that paradigm choice can fully account for our results, however. If anything, it could be argued that we underestimated proficiency effects because the cheating paradigm provided limited opportunities for elaboration (and subsequent cue/impression differences), but we still found effects. Moreover, no differences in proficiency effects were found when two comparable paradigms—lengthy alibis versus shorter opinion—were recently compared (Evans et al., 2017). Given the pattern of results, and the extant literature (e.g., Evans & Michael, 2014; Leach & Da Silva, 2013), it is much more likely that our effects were due to the speakers' levels of proficiency rather than paradigm.

Only two levels of proficiency were compared: nonnative and native English speakers. This methodological choice limited our ability to conduct a nuanced examination of proficiency. Evidence from two recent studies—in which four levels of proficiency were compared—suggests that there are primarily differences between beginner nonnative speakers and native speakers (i.e., groups that we tested; Elliott & Leach, 2016; Evans et al., 2017). Further research is still needed, however, given that those studies lacked systematic cue analyses and an examination of observers' impressions.

The number of interviewees in our cue analyses may also be a limitation. According to Hartwig and Bond (2011), variance in lie-detection accuracy may be caused by the existence of only small differences between lie- and truth-tellers. Thus, it is important to examine interviewees' cues closely to understand shortcomings in observer accuracy. In line with previous research (e.g., Evans et al., 2013), we analysed the same interviewees ($n = 30$) about which observers made accuracy decisions. Furthermore, we were limited by the available stimuli set because we utilized stimuli collected for a previous study (see Da Silva & Leach, 2013). A more robust future examination of cues, however, should utilize a greater number of interviewees per cell. In addition, we limited the scope of our statistical analyses to significant cues to deception and a handful of behaviours that previous research had suggested could be affected by proficiency (e.g., DePaulo et al., 2003). The list was by no means exhaustive. Future studies might incorporate additional items, particularly linguistic cues that are likely to vary between native and nonnative speakers. Researchers could also examine additional impression items to obtain a more thorough account of how deceptive and truthful nonnative speakers are perceived by observers.

Another potential limitation is our use of multiple raters who coded without entirely overlapping. In a recent article, Sporer et al. (2021) recommended that all stimuli should be double rated. However, in the present study, our raters' cue coding only overlapped in a proportion of all ratings (i.e., 25%). We established this guideline prior to the commencement of coding based on the best practices outlined in the Datavyu's manual (Datavyu Team, 2014). Double rating should become the standard for coding stimuli in future studies.

Finally, our conservative approach (i.e., adopting stringent Bonferroni-adjusted α levels) may have led us to overlook patterns in the data. If we had used the standard significance level (i.e., .05), then we would have also found that nonnative speakers and lie-tellers were perceived less positively (e.g., paused longer, made more negative statements) than native speakers who were either lying or telling the truth (e.g., appeared believable, capable, confident, intelligent) in both cues and impressions analyses.

Implications and Conclusions

It is important to evaluate how the growing number of nonnative speakers worldwide would be categorized and perceived, particularly in forensic contexts. Our findings indicate that nonnative speakers are viewed more negatively, and exhibit different behavioural profiles, than native speakers. Perhaps as a result, observers cannot detect their deception. These findings are problematic for several reasons. They undermine the legal tenet that all—including those with different language proficiencies—should be equal under the law. In addition, they might interfere with the integrity of the justice system: A proportion of guilty nonnative speakers may be overlooked or freed, and a proportion of innocent nonnative speakers may be wrongly accused or convicted of a crime. Thus, nonnative speakers appear to be at a significant disadvantage in lie-detection contexts.

Résumé

Nous avons examiné la façon dont les impressions d'observateurs au sujet de locuteurs non natifs et leurs indices de tromperie a influé sur leur prise de décisions. Les locuteurs anglophones natifs et non natifs ont menti ou dit la vérité au sujet d'avoir commis une transgression, et les observateurs essayaient de déceler leurs mensonges. Les observateurs ont mieux réussi à faire la distinction entre les mensonges et les vérités exprimés par les locuteurs natifs que par les locuteurs non natifs. Ils ont aussi eu plus d'impressions positives au sujet des locuteurs natifs qu'au sujet des locuteurs non natifs. Contrairement aux observateurs, les encodeurs ont repéré de multiples différences parmi les présentations des indices de tromperie des personnes interviewées chez les groupes aux niveaux de compétence variés. Dans l'ensemble, les locuteurs non natifs semblent désavantagés dans les contextes de détection de mensonges.

Mots-clés : détection de mensonges, prise de décisions, compétence linguistique, indices de mensonges, impressions

References

Akehurst, L., Arnhold, A., Figueiredo, I., Turtle, S., & Leach, A.-M. (2018). Investigating deception in second language speakers: Interviewee and

assessor perspectives. *Legal and Criminological Psychology*, *23*(2), 230–251. https://doi.org/10.1111/lcrp.12127

Ardila, A. (2003). Language representation and working memory with bilinguals. *Journal of Communication Disorders*, *36*(3), 233–240. https://doi.org/10.1016/S0021-9924(03)00022-4

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*(11), 417–423. https://doi.org/10.1016/S1364-6613(00)01538-2

Blau, E. K. (1991). *More on comprehensible input: The effect of pauses and hesitation markers on listening comprehension* [Paper presentation]. Annual Meeting of the Teachers of English to Speakers of Other Languages in San Juan, Puerto Rico.

Bond, C. F., Jr., & Atoum, A. O. (2000). International deception. *Personality and Social Psychology Bulletin*, *26*(3), 385–395. https://doi.org/10.1177/0146167200265010

Bond, C. F., Jr., & DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality and Social Psychology Review*, *10*(3), 214–234. https://doi.org/10.1207/s15327957pspr1003_2

Broadbent, D. E. (1957). A mechanical model for human attention and immediate memory. *Psychological Review*, *64*(3), 205–215. https://doi.org/10.1037/h0047313

Canadian Language Benchmarks. (2012). *Canadian language benchmarks: English as a second language for adults*. http://www.cic.gc.ca/english/pdf/pub/language-benchmarks.pdf

Castillo, P. A., Tyson, G., & Mallard, D. (2014). An investigation of accuracy and bias in cross-cultural lie detection. *Applied Psychology in Criminal Justice*, *10*(1), 66–82. http://dev.cjcenter.org/_files/apcj/apcjspring2014castillo.pdf_1399047724.pdf

Cheng, K. H. W., & Broadhurst, R. (2005). The detection of deception: The effects of first and second language on lie detection ability. *Psychiatry, Psychology, and Law*, *12*(1), 107–118. https://doi.org/10.1375/pplt.2005.12.1.107

Da Silva, C. S., & Leach, A. M. (2013). Detecting deception in second-language speakers. *Legal and Criminological Psychology*, *18*(1), 115–128. https://doi.org/10.1111/j.2044-8333.2011.02030.x

Datavyu Team. (2014). *Datavyu: A video coding tool*. Databrary Project. http://datavyu.org

DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *American Psychological Association*, *129*(1), 74–118. https://doi.org/10.1037/0033-2909.129.1.74

Elliott, E., & Leach, A. M. (2016). You must be lying because I don't understand you: Language proficiency and lie detection. *Journal of Experimental Psychology: Applied*, *22*(4), 488–499. https://doi.org/10.1037/xap0000102

Evans, J. R., & Michael, S. W. (2014). Detecting deception in non-native English speakers. *Applied Cognitive Psychology*, *28*(2), 226–237. https://doi.org/10.1002/acp.2990

Evans, J. R., Michael, S. W., Meissner, C. A., & Brandon, S. E. (2013). Validating a new assessment method for deception detection: Introducing a psychologically based credibility assessment tool. *Journal of Applied Research in Memory and Cognition*, *2*(1), 33–41. https://doi.org/10.1016/j.jarmac.2013.02.002

Evans, J. R., Pimentel, P. S., Pena, M. M., & Michael, S. W. (2017). The ability to detect false statements as a function of the type of statement and the language proficiency of the statement provider. *Psychology, Public Policy, and Law*, *23*(3), 290–300. https://doi.org/10.1037/law0000127

Frayer, J. M., & Krasinski, E. (1995). Perception of hesitation in nonnative speech. *Bilingual Review/La Revista Bilingue*, *20*(2), 114–121. https://www.jstor.org/stable/25745268

Gill, M. M. (1994). Accent and stereotype: Their effect on perceptions of teachers and lecture comprehension. *Journal of Applied Communication Research*, *22*(4), 248–361. https://doi.org/10.1080/00909889409365409

Global Deception Research Team. (2006). A world of lies. *Journal of Cross-Cultural Psychology*, *37*(1), 60–74. https://doi.org/10.1177/0022022105282295

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. Wiley.

Gregersen, T. S. (2005). Nonverbal cues: Clues to the detection of foreign language anxiety. *Foreign Language Annals*, *38*(3), 388–400. https://doi.org/10.1111/j.1944-9720.2005.tb02225.x

Hartwig, M., & Bond, C. F., Jr. (2011). Why do lie-catchers fail? A lens model meta-analysis of human lie judgments. *Psychological Bulletin*, *137*(4), 643–659. https://doi.org/10.1037/a0023589

Hosman, L. A., & Wright, J. W., II. (1987). The effects of hedges and hesitations on impression formation in a simulated courtroom context. *Western Journal of Speech Communication*, *51*(2), 173–188. https://doi.org/10.1080/10570318709374263

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Leach, A. M., & Da Silva, C. S. (2013). Language proficiency and police officers' lie detection performance. *Journal of Police and Criminal Psychology*, *28*(1), 48–53. https://doi.org/10.1007/s11896-012-9109-3

Leach, A. M., Snellings, R. L., & Gazaille, M. (2017). Observers' language proficiencies and the detection of non-native speakers' deception. *Applied Cognitive Psychology*, *31*(2), 247–257. https://doi.org/10.1002/acp.3322

Leach, A.-M., Da Silva, C. S., Connors, C. J., Vrantsidis, M. R. T., Meissner, C. A., & Kassin, S. M. (2019). Looks like a liar? Beliefs about native and non-native speakers' deception. *Applied Cognitive Psychology*, *34*(2), 387–396. https://doi.org/10.1002/acp.3624

Leary, M. R., & Kowalski, R. M. (1990). Impression management: A literature review and two-component model. *Psychological Bulletin*, *107*(1), 34–47. https://doi.org/10.1037/0033-2909.107.1.34

Lev-Ari, S., & Keysar, B. (2010). Why don't we believe non-native speakers? The influence of accent on credibility. *Journal of Experimental Social Psychology*, *46*(6), 1093–1096. https://doi.org/10.1016/j.jesp.2010.05.025

Lev-Ari, S., & Keysar, B. (2012). Less-detailed representation of non-native language: Why non-native speakers' stories seem more vague. *Discourse Processes*, *49*(7), 523–538. https://doi.org/10.1080/0163853X.2012.698493

Li, P., Sepanski, S., & Zhao, X. (2006). Language history questionnaire: A web-based interface for bilingual research. *Behavior Research Methods*, *38*(2), 202–210. https://doi.org/10.3758/BF03192770

Maclay, H., & Osgood, C. E. (1959). Hesitation phenomena in spontaneous English speech. *Word*, *15*(1), 19–44. https://doi.org/10.1080/00437956.1959.11659682

Rinne, J. O., Tommola, J., Laine, M., Krause, B. J., Schmidt, D., Kaasinen, V., Teräs, M., Sipilä, H., & Sunnari, M. (2000). The translating brain: Cerebral activation patterns during simultaneous interpreting. *Neuroscience Letters*, *294*(2), 85–88. https://doi.org/10.1016/S0304-3940(00)01540-8

Russano, M. B., Meissner, C. A., Narchet, F. M., & Kassin, S. M. (2015). Investigating true and false confession within a novel experimental paradigm. *Psychological Science*, *16*(6), 481–486. https://web.williams.edu/Psychology/Faculty/Kassin/files/Russano_et_al_05.pdf

Sert, O., & Jacknick, C. M. (2015). Student smiles and the negotiation of epistemics in L2 classrooms. *Journal of Pragmatics*, *77*, 97–112. https://doi.org/10.1016/j.pragma.2015.01.001

Spence, S. A., Farrow, T. F. D., Herford, A. E., Wilkinson, I. D., Zheng, Y., & Woodruff, P. W. R. (2001). Behavioural and functional anatomical correlates of deception in humans. *Neuroreport*, *12*(13), 2849–2853. https://doi.org/10.1097/00001756-200109170-00019

Sporer, S. L., Manzanero, A. L., & Masip, J. (2021). Optimizing CBCA and RM research: Recommendations for analyzing and reporting data on

content cues to deception. *Psychology, Crime & Law*, *27*(1), 1–39. https://doi.org/10.1080/1068316X.2020.1757097

Sporer, S. L., & Schwandt, B. (2006). Paraverbal indicators of deception: A meta-analytic synthesis. *Applied Cognitive Psychology*, *20*(4), 421–446. https://doi.org/10.1002/acp.1190

Sporer, S. L., & Schwandt, B. (2007). Moderators of nonverbal indicators of deception: A meta-analytic synthesis. *Psychology, Public Policy, and Law*, *13*(1), 1–34. https://doi.org/10.1037/1076-8971.13.1.1

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, *31*(1), 137–149. https://doi.org/10.3758/BF03207704

Suchotzki, K., & Gamer, M. (2018). The language of lies: Behavioral and autonomic costs of lying in a native compared to a foreign language. *Journal of Experimental Psychology: General*, *147*(5), 734–746. https://doi.org/10.1037/xge0000437

Tavakoli, P., & Foster, P. (2011). Task design and second language performance: The effect of narrative type on learner output. *Language Learning*, *61*(2), 439–473. https://doi.org/10.1111/j.1467-9922.2008.00446.x

Temple, L. (1992). Disfluencies in learner speech. *Australian Review of Applied Linguistics*, *15*(2), 29–44. https://doi.org/10.1075/aral.15.2.03tem

van 't Veer, A. E., Gallucci, M., Stel, M., & van Beest, I. (2015). Unconscious deception detection measured by finger skin temperature and indirect veracity judgments-results of a registered report. *Frontiers in Psychology*, *6*, Article 672. https://doi.org/10.3389/fpsyg.2015.00672

Volz, S., Reinhard, M.-A., & Müller, P. (2020). Why don't you believe me? Detecting deception in messages written by nonnative and native speakers. *Applied Cognitive Psychology*, *34*(1), 256–269. https://doi.org/10.1002/acp.3615

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

Vrij, A., Fisher, R., Mann, S., & Leal, S. (2008). A cognitive load approach to lie detection. *Journal of Investigative Psychology and Offender Profiling*, *5*(1–2), 39–43. https://doi.org/10.1002/jip.82

Vrij, A., Leal, S., Mann, S., Warmelink, L., Granhag, P. A., & Fisher, R. (2010). Drawings as an innovative and successful lie detection tool. *Applied Cognitive Psychology*, *24*(4), 587–594. https://doi.org/10.1002/acp.1627

Vrij, A., Mann, S. A., Fisher, R. P., Leal, S., Milne, R., & Bull, R. (2008). Increasing cognitive load to facilitate lie detection: The benefit of recalling an event in reverse order. *Law and Human Behavior*, *32*(3), 253–265. https://doi.org/10.1007/s10979-007-9103-y

Vrij, A., & Winkel, F. W. (1991). Cultural patterns in Dutch and Surinam nonverbal behaviour: An analysis of simulated police/citizen encounters. *Journal of Nonverbal Behavior*, *15*(3), 169–184. https://doi.org/10.1007/BF01672219

Vrij, A., & Winkel, F. W. (1994). Perceptual distortions in cross-cultural interrogations: The impact of skin color, accent, speech style, and spoken fluency on impression formation. *Journal of Cross-Cultural Psychology*, *25*(2), 284–295. https://doi.org/10.1177/0022022194252008

Walczyk, J. J., Roper, K., Seemann, E., & Humphrey, A. M. (2003). Cognitive mechanisms underlying lying to questions: Response time as a cue to deception. *Applied Cognitive Psychology*, *17*(7), 755–774. https://doi.org/10.1002/acp.914

Zuckerman, M., DePaulo, B. M., & Rosenthal, R. (1981). Verbal and nonverbal communication of deception. *Advances in Experimental Social Psychology*, *14*, 1–59. https://doi.org/10.1016/S0065-2601(08)60369-X

Received November 10, 2021
Revision received March 30, 2022
Accepted April 4, 2022 ■