In press, *Psychiatry, Psychology, and Law*

Clusters of Nonverbal Behavior Differentiate Truths and Lies about Future Malicious Intent in Checkpoint Screening Interviews

David Matsumoto and Hyisung C. Hwang

San Francisco State University and Humintell

Author Information

David Matsumoto (Corresponding Author)
Department of Psychology
San Francisco State University
1600 Holloway Avenue
San Francisco CA 94132
Email: dm@sfsu.edu
TEL: 415-338-1114
ORCID ID: 0000-0002-3186-1503

Hyisung C. Hwang
Humintell
PO Box 1304
El Cerrito, CA 94530
Email: hshwang@humintell.com
TEL: 510-620-9668

Word Count: 7,685

Abstract

Recent research has shown that nonverbal behavior (NVB) assessed across multiple channels can differentiate truthtellers from liars. No study, however, has examined whether multiple NVB can differentiate truths from lies about intent for future malicious behavior, or across multiple culture/ethnic groups. We address this gap by examining truths and lies about intent to commit a malicious act in the future in brief, checkpoint-type security screening interviews. Data from four NVB channels producing 21 observable NVB were coded and analyzed using different analytic strategies. Clusters of NVB differentiated truthtellers and liars at statistically significant levels and substantially beyond that by human observers. These findings showed that clusters of NVB can differentiate truthtellers from liars even in brief, checkpoint interviews.

Keywords:

Nonverbal Behavior, Veracity, Deception, Intent, Facial Expressions, Gestures, Voice, Body Movements, Fear, Disgust, Shrugs

Clusters of Nonverbal Behavior Differentiate Truths and Lies about Future Malicious Intent in Checkpoint Screening Interviews

Research has examined nonverbal behavior (NVB) associated with veracity and deception for decades with good reason: Findings have theoretical implications concerning cognitions and emotions associated with truthtelling and lying and practical ramifications to applications in real-life investigative settings. Individuals and organizations from many such settings, including law enforcement, national security, asset protection, and the business world have interests in utilizing reliable behavioral indicators of veracity, deception, and other mental states in interviews and investigations.

Almost two decades ago, a seminal meta-analysis concluded that few NVB differentiated truthtellers from liars in low stakes lies or when suppressing emotions or pain (DePaulo et al., 2003). Subsequent research, however, has examined multiple rather than single NVB and has provided evidence that NVB clusters do so. A meta-analysis examining NVB clusters reported that lies were detected with nearly 68% accuracy across settings (Hartwig & Bond, 2014), and concluded, “The higher accuracy rates obtained here suggest that signals of deception are manifested in constellations rather than single cues” (p. 667).[1] Other studies not included in that meta-analysis have also provided such evidence (Davis, Markus, Walters, Vorus, & Connors, 2005; Dunbar, Jensen, Tower, & Burgoon, 2014; Duran, Dale, Kello, Street, & Richardson, 2013; Ekman, O'Sullivan, Friesen, & Scherer, 1991; Jensen, Meservy, Burgoon, & Nunamaker, 2010; Matsumoto & Hwang, 2018a; Pennebaker & Chew, 1985; Vrij, Edward, Roberts, & Bull, 2000; Wright Whelan, Wagstaff, & Wheatcroft, 2014).

That multiple rather than single NVB better differentiate veracity and deception makes theoretical sense. Communication involves multiple NVB channels, all of which produce multiple signals that convey multiple messages. These messages include specific emotions, general affective orientations, words or phrases, general cognitive processes, physical effort, attitudes, or conversation regulation. In interaction, signals of any of these messages may be produced across any channel, with or without words, rendering the NVB system the most complex communicative system of the body.

NVB system complexity is compounded by the complexity of truthtellers' and liars' minds and associated cognitive and emotional loads (Ekman, 1985; Frank, 2009; Vrij, 2008). Liars lie about their thinking and think about their lying, and need to lie about their emotions and have emotions about the fact they are lying. All this occurs while attempting to manage impressions of themselves by others and maintaining cognitive and emotional control (Buller & Burgoon, 1996; Hurley & Frank, 2011; Vrij, 2008).

Because most people verbalize only a portion of their mental contents, and because different mental states map onto different NVB channels, truthful vs deceptive signaling can occur in multiple channels and in different ways. Emotionally, a fear of being caught, or guilt or even thrill about lying, may be betrayed by face or voice; nervousness may be betrayed by voice and body. Cognitively, obfuscation, fabrication, or omission in words may be betrayed by voice and gesture; lies about emotions and sensations experienced about an incident can range from anger or fear to shame or glee (cf, see the Reality Monitoring perspective; Johnson, 1988; Johnson & Raye, 1981). A myriad of possibilities in cognitions and emotions combined with complexity of the NVB system to signal them suggest that analysis of single NVB channels will underestimate the potential for the NVB *system* to differentiate truthtellers and liars, but analysis of NVB clusters across multiple channels may be better, which is what has been evidenced.

Within the deception literature, most studies have examined lies concerning an incident in the past. Some have examined lies about future intent (Granhag & Mac Giolla, 2014), but only few were related to intent for future malfeasance in criminal contexts (Matsumoto & Hwang, 2018b; Matsumoto, Hwang, & Sandoval, 2015; Vrij, Leal, Mann, & Granhag, 2011).[2] Theoretically, such lies may be different than lies about the past because they may access different domains of cognition and memory (Addis, Wong, & Schacter, 2007; Schacter & Addis, 2007; Schacter, Addis, & Buckner, 2008). Examining lies about future intent has practical implications; results of such studies may inform practitioners about behavioral indicators that occur in brief, checkpoint-type interactions, with implications to security procedures in settings that assess future malicious intent.

But examining lies about future intent in encounters such as checkpoint interviews is risky because they are brief and questions are not directed to specific incidents. Thus, they may not produce cognitive and emotional loads that differentiate truthtellers and liars as do longer, investigative interviews. Moreover, NVB tend to produce weak signals, as evidenced by effect sizes reported for single NVB in previous meta-analyses (DePaulo et al., 2003; Hartwig & Bond, 2014).

Also, most studies have involved participants in single cultures, and the complexity of the NVB system is complicated by the existence of cultural similarities and differences in the use of NVB. Although certain facial expressions of emotion are universally expressed and recognized, the contexts, dynamics, and rules that govern their usage and modification are culturally different (Hwang & Matsumoto, 2016a). While speech illustrative gestures function similarly across cultures, their form, frequency, and amplitude are culturally different, as are emblematic gestures (Cartmill & Goldin-Meadow, 2016). Body movements, including postures, gait, proxemics, and haptics also involve both cross-culturally similar and different aspects (Matsumoto, Hwang, & Frank, 2016). Vocalics appear to be cross-culturally similar in their characteristics to signal emotion but culturally different in signals of cognition, and speech articulation (Scott & McGettigan, 2016).

Overview of the Current Study

This study addressed the possibility of multiple NVB to differentiate truths from lies about future malicious intent in a culturally diverse community sample that participated in a mock crime experiment about a future theft. Stakes were associated with performances and manipulation checks ensured that participants perceived the stakes at moderate-high levels and were emotionally aroused. Participants were interviewed in a checkpoint-type screening interview prior to their gaining access to an area where a theft could occur. The context, therefore, was analogous to real-life settings in which individuals with malicious intent need to hide their intentions, with stakes involved regarding whether or not they were believed.

Participants' behavior was audio-video recorded using two cameras, one with straight head and shoulder shots and a second with a 90-degree angle of their whole bodies. We extracted data from four NVB channels – facial expressions of emotion (seven types), gestures (three types), whole body movements (four types), and vocalics (seven types). These NVB were selected a priori because previous research has tested their ability to differentiate truthtellers from liars (see review in DePaulo et al., 2003; Ekman et al., 1991; Matsumoto & Hwang, 2018a, 2018b), but mostly when tested individually and with some contradictory findings. The current study is the only to test NVB clusters in an initial screening interview about future intent.

We tested whether NVB differentiated truthtellers from liars singly (univariate analyses) and in combination (multivariate analyses). We included univariate analyses because each NVB signal itself should be relatively weak, and practitioners in real life would not be observing as many channels and signals as we assessed and could not perform complex mathematical algorithms in real time that account for interdependencies among observed variables as multivariate statistics would. We used two criteria in the univariate analyses to determine if NVB differentiated truthtellers from liars: null hypothesis significance testing (NHST) and effect sizes. NHST is dependent on sample sizes whereas effect sizes are less so and findings across studies can differ depending on sample size.

We hypothesized that clusters of NVB from different source channels would differentiate truthtellers from liars. Specifically, we predicted that at least one NVB from at least two channels would differentiate veracity and deception, as opposed to multiple NVB within the same channel (for example, facial expressions of anger and fear, which both originate in the same face channel). We further hypothesized that culture/ethnicity would moderate the findings.

Method[3]

Design

The experiment was a two-way design involving Veracity (truths vs. lies) and Culture/ethnicity (European Americans and Chinese, Hispanic, and Middle Eastern immigrants). Other papers have examined verbal and NVB coded from a different interview in this experiment (Matsumoto & Hwang, 2015, 2018a; Matsumoto et al., 2015), as well as timing characteristics of facial expressions from a subset of the records reported below (Matsumoto & Hwang, 2018b). The current study reports occurrence data (ignoring timing) of four sources of NVB from the initial interview of the study ($N = 226$), all of which have not been analyzed or reported previously and are new to the literature.[4]

Participants

Community samples were recruited in the San Francisco Bay Area and Buffalo, NY. European Americans were U.S. born-and-raised Caucasians; immigrant participants were born and raised in their home country or the U.S., their first language was of the home country, and both parents were born and raised in the home country. Home country and first language were as follows: Chinese: People's Republic of China, Hong Kong, or Taiwan and Mandarin or Cantonese; Hispanics: any country in Central or South America and Spanish; Middle East: any country in Northern Africa or Western Asia and Arabic. A total $N = 226$ included *n*s = 40 and 38, 46 and 36, 28 and 18, and 8 and 12, in the lie and truth conditions for the European American, Chinese, Hispanic, and Middle Eastern samples, respectively. Sex ratio was roughly evenly distributed (47.4% males, 52.6% females; and $M_{age} = 27.32$, range 19-47). The experiment was administered in English.[5]

Measures

Pre-session measures included a demographics questionnaire, the General Ethnicity Questionnaire (GEQ; Tsai, Ying, & Lee, 2000), the Machiavellianism Scale (Christie, 1970), the Self-Monitoring Scale (Snyder, 1974), and an emotion checklist that included 12 emotion words (guilt, fear, anger, embarrassment, worry, contempt, excitement, disgust, amusement, nervousness, surprise, and interest) rated on 9-point scales labeled 0, None, 4, Moderate Amount, and 8, Extremely Strong. This checklist was also administered at the end of the experiment as a manipulation check on emotional arousal.

The GEQ served as a manipulation check for ethnic/cultural differences (target group was made applicable for each group). The Chinese sample had significantly higher scores on the GEQ Total than Chinese who immigrated to the U.S. before age 12 and American born Chinese

(Tsai et al., 2000), $t(74) = 8.07$, $p < .001$, $d = .93$; and $t(74) = 1.71$, $p < .05$, $d = .20$, respectively. GEQ scores for Middle Eastern and Hispanic immigrants were comparable to the Chinese.

Procedures

After completing the pre-session measures, participants were informed that they would be randomly assigned to either steal a $100 check or not, go through up to three interviews, and earn a minimum of $20. If judged honest, participants were told that they would receive additional money and be allowed to leave early; if judged dishonest, they would receive no additional money and would have to stay longer to complete other procedures. Participants rated these stakes on a scale from 1, No consequence, even slightly pleasurable, to 10, Maximum consequence, even slightly painful. The mean was above the midpoint (5.68; SD = 2.24) and significantly greater than one, $t(224) = 31.34$, $p < .001$, $d = 2.09$; there were no ethnicity or condition differences. Participants were then randomly assigned to steal or not.

An initial screening interview occurred in an area modeled after a security checkpoint. An interviewer entered the area, went to a podium, and instructed the participant to go through a metal detector and step up to the podium. The screening interview included seven brief questions lasting an average of 1:56 m.[6] Interviewers were 10 males above age 30 trained to deliver interviews neutrally. (The interviewers were predominantly European American, with one African American and one Asian American; thus, counterbalancing race or ethnicity between interviewers and interviewees was not done.) When completed, the interviewer left and the remainder of the experiment proceeded, including a secondary interview, the mock crime, an investigative interview, administration of post-session measures, debriefing, post-session consent, and payment. All coding and analyses below focused on three questions in the initial screening interview to which liars had to lie but truthtellers could tell the truth; coding for each channel was performed by different sets of coders.

Coding

Facial expressions of emotion. Facial expressions of seven emotions (anger, contempt, disgust, fear, happiness, sadness, and surprise) were coded by two coders blind to veracity condition using Emotion FACS (EMFACS; Hwang & Matsumoto, 2016b; Matsumoto, Ekman, & Fridlund, 1991), an abbreviated version of the Facial Action Coding System (FACS; Ekman & Friesen, 1978). Reliabilities were computed on a subsample of records (n = 77) and were acceptable (% agreement = .70, .91, .75, .96, .67, .68, and .82 for anger, contempt, disgust, fear, happiness, sadness and surprise respectively; and for total emotions coded, $r(76) = .97$). Analyses used frequencies of each emotion for which both coders agreed across the target questions.

Gestures. Head nods, headshakes, and shrugs (including shoulder and face shrugs) were coded. Four raters coded 1/3rd of all videos and produced an average reliability of .76. Coders then coded all remaining videos. Mid-coding reliability between the coders and authors was .90 across 2/3rds of the total sample. Analyses used frequencies of each gesture across the target questions.

Whole body movements. Four whole body movements were coded by two coders using video from a second camera that was at a 90-degree angle to the interaction in which participants' bodies could be seen (available for only n = 139 interviews): Body Trembling, including trembling of body or voice; Body Swaying, involving rocking back and forth or side to side; Fidgeting/Grooming, including any rubbing or wringing of hands, face touching, shuffling or tapping of feet, or any kind of grooming behavior; and Rigid Posture, involving minimal, stiff body movements with arms close to the sides. Reliability estimates and coder arbitration and calibration occurred on n = 50 cases and were high and acceptable for all four movements, αs = .73, .95, .96, and .77, respectively. Both coders then coded all video records available. Analyses included scores for each movement averaged between the coders across the target questions.

Vocal Data

Vocal data were extracted using PRAAT, an open source software. The following variables were generated across the target questions: Pitch, Pitch Range, Intensity, Intensity Range, Response latency (generated from onset/offset interview logs), Duration and unfilled pauses (combined into one variable), and Speech and articulation rates (combined into one variable).

Interview Contamination

Two coders blind to conditions and hypotheses independently coded transcripts from 30 cases for participant understanding of the questions and interviewer contamination. Reliabilities were high for both (rs = .97 and .83, respectively). The remaining cases were then coded by one coder. Analyses below included only cases for which these codes were zero.[7]

Results

Manipulation Check

A Pre-post (2) x Emotion (12) x Veracity (2) mixed ANOVA was computed on the self-reported emotions. The Pre-post x Veracity and three-way interactions were significant, $F(1, 186) = 18.49$, $p = .000$, $\eta_p^2 = .09$; and $F(11, 2046) = 4.96$, $p = .000$, $\eta_p^2 = .03$. Pre-post x Veracity simple interaction contrasts produced significant effects on seven emotions. Liars reported increases in guilt, fear, embarrassment, worry, and nervousness; truthtellers reported less excitement and interest, $.02 < \eta_p^2 < .19$. Thus, participants were emotionally aroused and the emotions were elicited differentially in the truthtellers and liars.

Univariate Analyses

We computed descriptives (Ms and SDs) on all NVB variables and then two initial MANOVAs, one for whole-body NVB and another for all other NVB because of differences in sample sizes associated with the codes available. Both were significant, Wilk's λ =.94, $F(4, 118) = 2.45$, $p = .050$; and Wilk's λ =.82, $F(17, 171) = 2.27$, $p = .004$; respectively. We then computed separate, two-tailed *t*-tests on each NVB singly to examine whether each differed by veracity condition. NHST indicated that three NVB differentiated truthtellers and liars at $p \leq .05$: liars produced more facial expressions of fear, shrugs, and fidgeting/grooming. Four other NVB trended toward significance: truthtellers produced more facial expressions of happiness and surprise, and had greater response latencies and duration and unfilled pauses. (Because variables trending toward significance may have operational relevance and influence the multivariate effects, we report them.) The number of variables that differentiated truthtellers and liars was greater than that what would be expected by chance.[8]

We also used Cohen's $d \geq .20$ as an effect size criterion reduce reliance on NHST to identify NVB that differentiated truthtellers and liars. This criterion was chosen because it is typically interpreted as an estimate of "small" effects. We reckoned that any single NVB would only produce weak signals that differentiated truthtellers from liars, and wanted to utilize the minimal value that would also possibly be operationally relevant to observers. Relative to truthtellers, liars produced more facial expressions of fear, shrugs, fidgeting/grooming, and rigid postures while truthtellers produced more facial expressions of happiness and surprise, response latencies, and duration and unfilled pauses (see Table 1). Thus, regardless of the criterion used to identify variables, multiple NVB from different channels differentiated truthtellers from liars.

Examining for Possible Latent Structures

To examine if the NVB organized around latent factors, we computed principal components analyses on all variables. Kaiser criterion indicated nine factors accounting for 69.71% of the total variance. After Varimax rotation, we identified scales with variables with factor loadings ≥ .30 and computed Cronbach's alphas; with the exception of the first scale, all were low (αs = .91, .63, .57, .49, .22, .41, .14, .39, and .08). The scree plot was also inconclusive and linear (correlation between extracted factors and eigenvalues was $r(21) = -.97$). Analyses utilizing three-, four-, and five-factor solutions and two different oblique rotation methods also did not produce interpretable structures. We concluded that the NVB did not group together in an interpretable manner using factor techniques and proceeded with the analyses below using the variables separately.

Multivariate Analyses: Do Clusters of NVB Differentiate Truthtellers and Liars?

To determine if multiple NVB differentiated truthtellers from liars, we computed binary logistic regressions with backward conditional entry separately for the four NVB channels and then all NVB together. We selected final models based on the lowest *p*-values associated with the largest classification accuracy rates. If two models had the same classification accuracies, we selected the model with larger number of predictors.

The results are shown in Table 2. For facial expressions of emotion, fear and surprise were final predictors with an overall classification rate of 57.1%. For gestures, shrugs were a predictor with an overall classification rate of 60.1%. For whole body movements, the selected model trended toward significance with fidgeting/grooming a predictor and an overall classification rate of 57.7%. The selected model for voice also trended toward significance, with response latency a significant predictor and an overall classification rate of 60.2%.

When all NVB were analyzed together, five variables were significant predictors with an overall classification rate of 76.0%. Liars produced more facial expressions of disgust, shrugs, body swaying, and fidgeting/grooming, while truthtellers had larger pitch range. As predicted, these specific NVB came from multiple channels. We also report the final (i.e., most parsimonious) model in the analysis. Here, five variables emerged as predictors with an overall classification rate of 65.9%. Liars produced more facial expressions of disgust, shrugs, and fidgeting/grooming. Still, the final NVB came from multiple channels, not a single channel.

Thus, analyses of single channels of NVB produced classification rates that were only slightly above the average accuracy rate of human observers in distinguishing truths and lies (54% as reported by Bond & DePaulo, 2006). Analyzing NVB clusters across channels, however, produced substantially higher classification accuracy rates with NVB from multiple channels contributing to the predictions.

Possible Culture/Ethnicity Moderation

To test for culture/ethnicity moderation of the findings, we recomputed the analyses above using Culture/Ethnicity (4), Gender (2), and Veracity (2) MANOVAs using the 21 NVB as dependents, as we reckoned that culture/ethnicity moderation of the association between the NVB with veracity condition would qualify interpretations of both sets of findings above. The Culture/Ethnicity by Veracity condition as not significant, Wilk's λ =.76, $F(51, 468.22) = 1.01$, $p = .455$, suggesting that culture/ethnicity did not moderate the findings above.

Additional Analyses

Protection against Type I error. We recomputed the multivariate analyses using random data that were randomized with different ranges for different NVB given the actual range of values that occurred in the original data set. Univariate t-tests produced only one significant finding at $p \leq .05$ (as would be expected for 21 tests); truthtellers produced *more* shrugs, $t(225) = 2.11$, $p = .036$, which was contrary to what would be predicted. Another variable trended toward significance; liars had shorter duration and unfilled pauses, $t(225) = 1.82$, $p = .070$. At either criterion, the number of significant tests was not larger than what was expected by chance. The same two variables were the only variables to meet the Cohen's *d* criterion.

Log regressions using the random data and the same criteria as above produced a selected model that was significant, $\chi^2(4, 227) = 11.03$, $p = .026$ with an overall classification accuracy rate of 63.0%. The final variables in the selected model were shrugs (direction opposite to that reported in main analyses) and duration and unfilled pauses. The final model was also significant, $\chi^2(2, 227) = 7.26$, $p = .027$ with an overall classification accuracy rate of 57.7%. The final variables in the model were once again shrugs (opposite direction) and duration and unfilled pauses.

Thus, using random data, the number of significant univariate results was not different from that expected by chance, the direction of the findings would not have been predicted, and classification accuracies of the selected and final models in multivariate analyses were lower than that produced with actual data and in directions that would not have been predicted.

Possible gender moderation. The Gender by Veracity interaction from the overall Culture/Ethnicity by Gender by Veracity MANOVA reported above was not significant, Wilk's $\lambda = .91$, $F(17, 157) = .91$, $p = .561$, indicating that gender did not moderate the findings reported above. The gender main effect, however, was significant, Wilk's $\lambda = .34$, $F(17, 157) = 18.32$, *p*

< .001. Follow-up analyses indicated that gender differences existed on head nods, body trembling, pitch, and pitch range, $F(1, 177) = 6.04$, $p = .015$, $\eta_p^2 = .033$; $F(1, 107) = 7.59$, $p = .007$, $\eta_p^2 = .066$; $F(1, 175) = 233.07$, $p \leq .001$, $\eta_p^2 = .571$; and $F(1, 175) = 82.24$, $p \leq .001$, $\eta_p^2 = .320$, respectively. Females produced more head nods ($M = 5.14$, $SD = 4.72$) than males ($M = 3.31$, $SD = 3.56$) and higher pitch and pitch range ($M = 181.97$, $SD = 27.31$; $M = 167.38$, $SD = 73.77$) than males ($M = 118.13$, $SD = 16.58$; $M = 78.05$, $SD = 37.47$). Males produced more body trembling ($M = 1.48$, $SD = 3.10$) than females ($M = .52$, $SD = 1.08$).

Ethnicity main effects. The Culture/Ethnicity main effect from the same overall MANOVA reported above was significant, Wilk's λ =.35, $F(51, 468.22) = 3.93$, p <, .001. Follow-up tests were significant for facial expressions of anger, body swaying, pitch range, intensity, response latency, and duration and unfilled pauses. We followed these effects using pairwise comparisons with Scheffe corrections (Table 3). Chinese immigrants produced more facial expressions of anger than Hispanic immigrants and more body swaying and duration and unfilled pauses than Hispanic immigrants and European Americans. Middle Eastern immigrants and Hispanic immigrants had higher pitch range than European Americans and greater vocal intensity than Chinese immigrants.

Discussion

The findings added to a growing literature documenting that multiple NVB in clusters can differentiate truthtellers from liars by documenting this effect in lies about future malicious intent in brief, checkpoint interviews. Across analyses, liars produced more facial expressions of fear and disgust, shrugs, fidgeting/grooming, body swaying, and rigid posture than truthtellers while truthtellers produced more facial expressions of happiness and surprise, longer response latencies, greater pitch range, and longer durations and unfilled pauses. Clusters of NVB across multiple channels outperformed NVB in single channels, supporting contentions that NVB represent a complex and comprehensive bodily communication system that conveys multiple messages in multiple signals in multiple channels.

The type of lie tested was unique – the intent to commit a malicious act in the future. As mentioned above, many studies have already demonstrated that NVB clusters can differentiate truths from lies about the past (summarized in Hartwig & Bond, 2014). Lies about future intentions have been considered potentially different than lies about the past because they may recruit different cognitions. But the NVB that differentiated truthtellers and liars here were similar to those reported in previous studies, and were comparable in detection accuracy rates (68% according to the meta-analysis cited above compared to 76.0% and 65.9% found here). The current findings, therefore, suggested that the cognitive and emotional processes associated with veracity and deception may be similar regardless of whether about the past or future. Both involve lying about one's thinking and thinking that one is lying, and lying about one's emotions as well as having emotions about lying; produced NVB may be signals of this process. This interpretation is consistent with neuroimaging studies demonstrating similarities between remembering the past and imagining the future (Schacter et al., 2008; Schacter et al., 2012). Understanding cognitive and emotional processes underlying the act of lying may give clues about differences in mental complexities between truthtellers and liars that may be consistent across types of lies and produce similarities in NVB, which is what we observed.[9]

The specific NVB that emerged as indicators provided glimpses into mindset content differences between truthtellers and liars, and supported different theories about truthtellers and liars, including theories about emotional leakage (Ekman, 2009), impression management (Buller & Burgoon, 1996; Burgoon & Buller, 1994; DePaulo et al., 2003), cognitive load (Vrij, 2008), emotional control (Frank & Svetieva, 2013), or a combination of these (Zuckerman, DePaulo, & Rosenthal, 1981). For example, that liars produced more signals of fear and disgust suggested that appraisals occurred related to threat or contamination, related to leakage. Liars produced more shrugs, suggestive of doubt or uncertainty, related to cognitive load. Their greater body swaying and fidgeting/grooming were suggestive of more overall nervousness (leakage) and their smaller pitch ranges were suggestive of greater vocal control (control or impression management). Liars also displayed less happiness and surprise, indicative of an overall negative, tense, uncertain state of mind (leakage). Thus, examination of NVB clusters opens the door to the possibility that multiple theories of deception may be supported because different NVB may signal different content related to those theories.

Findings were different across analyses. Facial expressions of happiness, shrugs, and fidgeting/grooming emerged as indicators in both univariate and multivariate analyses. Other NVB – facial expressions of fear, disgust, and surprise; and rigid posture, response latency, duration and unfilled responses, body swaying, and pitch range – emerged in one or the other analysis. These differences likely occurred because of the way NVB were handled as dependents. Multivariates created linear combinations of NVB by weighting the variables depending on intercorrelations conditional on optimizing group classification; the univariates did not. Although multivariates produced the most statistically parsimonious combination of variables relevant to category prediction, these computational differences also meant that different findings emerged depending on the analyses conducted and criteria adopted.

Assessing NVB clusters raises methodological concerns, especially about Type I error. This problem is compounded when sample sizes do not allow for sufficient power and findings are optimized for individual data sets, which lead to concerns about replicability and generalizability (for an excellent discussion of these issues see Luke, 2019). We attempted to mitigate this problem through re-analyses using random data, which did not produce a number of significant findings beyond that expected by chance, and those produced were contradictory to what would have been expected.

Another way to mitigate Type I error is to examine consistencies (and inconsistencies) in findings across studies that examined clusters. The only other comparable study did so in an investigative interview in the same experiment (Matsumoto & Hwang, 2018a). That interview was longer than the screening interview in this study, included different types of questions, and analyses were computed separately for question types. During open-ended questions, liars showed fewer facial expressions of anger and happiness and more disgust, fear, and surprise. Liars also had fewer head nods; lower voice pitch, intensity, and duration and unfilled pauses; and greater pitch range and intensity range. During indicator questions, liars produced more anger and disgust and less head nods. Thus, consistent with our findings, truthtellers and liars differed on facial expressions of happiness, disgust, and fear. Contradictory findings occurred on facial expressions of surprise, pitch range and duration and unfilled pauses. And several findings occurred in one study but not the other: shrugs and response latency emerged in this study but not the previous while facial expressions of anger, head nods, pitch, intensity, and intensity range emerged in the previous but not the current. Fidgeting/grooming, rigid posture, and body swaying were not measured in previous study.

These differences may have occurred for several reasons, one being differences in context. The current study involved brief, checkpoint interviews with short questions asking about future intent. The previous study involved longer interviews with open ended and indicator questions designed to accentuate differences between truths and lies about a past event. Such context differences (in type and object of lie, investigative context, and nature of questions) likely contributed to differences between the findings. For example, in this study, truthtellers had longer response latencies, greater pitch range, and longer durations and unfilled pauses than liars, which was the opposite reported in Matsumoto and Hwang (2018a). These differences probably occurred because of the length of interviews and types of questions asked. To wit, question type analyses in the previous study indicated that direct questions (like those in the current study) produced less NVB in general. Longer, open ended questions may have allowed for greater complexity of mental states resulting in greater NVB production, thus increasing the range of NVB that differentiated truthtellers and liars.

But another possibility is that some findings are not replicable. Future reviews should compare studies that examined NVB clusters to identify patterns of consistent NVB that emerge. Unfortunately, such a volume of research does not yet exist and we hope that findings from this study and others can serve as a call for such research. Doing so may require some consistency in selection and measurement of NVB across studies and investigators, which rarely occurs, rendering integrative, synthetic reviews difficult. Researchers may need to consider which NVB to include in a more comprehensive and systematic assessment in the future.

That neither culture/ethnicity (or gender) moderated the findings was also interesting, suggesting the potential cross-cultural/ethnic utility of NVB analyses. Although there is ample evidence for cultural/ethnic differences in NVB production (see reviews in Cartmill & Goldin-Meadow, 2016; Hwang & Matsumoto, 2016a; Matsumoto et al., 2016; Scott & McGettigan, 2016), the checkpoint situation with brief questions may have reduced the possibility for such differences to emerge because of a focus on quick reactions and responses. Cultural/ethnic differences may occur with longer conversations, giving interviewees time for cultural factors to influence responses. Although there is evidence for liars to take longer to respond and with less verbal volume (Newman, Pennebaker, Berry, & Richards, 2003; Vrij, 2008), in the current study truthtellers had longer response latencies and durations, and there were no differences in speech or articulation rates.

There were, however, culture/ethnic differences main effects in overall usage of facial expressions of anger, body swaying, pitch range, vocal intensity, response latency, and duration and unfilled pauses when in interaction. These differences pointed to the very real possibility that these NVB be mistakenly interpreted as deception indicators. Interpretation of these findings, however, also need to be tempered because of the sample sizes used in the analyses, which is a limitation of the study. All findings should be replicated in the future with larger sample sizes and other culture/ethnic groups, especially those in other nation/cultures with larger cultural distances.

Another limitation of the study was that it included no a priori predictions. Future theoretical and empirical work should address that and would need to include theories about individual differences in emotions and cognitions when truthtelling and lying that would lend themselves to predictions of which specific NVB would be produced when lying. Hopefully, such studies can involve pre-registered hypotheses.

The current findings suggested that practitioners be aware of multiple channels of NVB when conducting credibility assessments. Our findings may be especially meaningful for practitioners who interact with people of different cultural/ethnic backgrounds to assess credibility when only limited information is available moment to moment. In these cases, discrete, nonverbal signals can be potential landmarks of where to explore and in priority order. Even when language is a barrier and resources are limited, interactions do not have to be prohibited, but a focus can be altered to non-linguistic approaches. Vigilant observation of multiple NVB channels in interaction is challenging because NVB signals, especially when observed singly, are weak. One way to mitigate the challenge of leveraging weak signals in practice is to focus on validated indicators, which would provide practitioners with additional insights concerning the interviewees' mindsets and be a strength for interviewers.

Another way to streamline efficiency for practitioners is to focus on relatively smaller clusters of NVB because different NVB have different relative contributions to the prediction of veracity and deception (Hartwig & Bond, 2014). In our study, this can be estimated using regression coefficients, which would give priority to facial expressions of emotions, shrugs, fidgeting/grooming, body swaying, and pitch range, in that order. Future research can examine whether this or any other NVB cluster may prove to be useful diagnostically; they may also be helpful in informing practitioners about validated behavioral indicators that occur in brief, checkpoint-type interactions, with potential implications to security procedures in a wide variety of settings that assess malicious intent.

Acknowledgement

This work was funded by the Transportation Security Administration (TSA) contract HSTS04-16-C-CT9002. Statements of fact, opinion and analysis in the paper are those of the authors and do not reflect the official policy or position of the TSA or the U.S. Government.

Ethical Standards

Declaration of Conflicts of Interests

Both authors are employees of Humintell, a for-profit company that conducts applied research and training in topics covered in this paper.

Ethical Approval

All procedures performed in studies involving human participants were in accordance with the ethical standards of the University at Buffalo, State University of New York Social and Behavioral Sciences Institutional Review Board.

Informed Consent

Written informed consent was obtained from all interviewees appearing in the video records after they were fully informed about the experimental procedures but before the interviews and once again after the experiment was completed.

Biographical Note

David Matsumoto is Professor of Psychology at San Francisco State University (San Francisco, CA) and Director at Humintell (El Cerrito, CA). Hyisung C. Hwang is Senior Research Scientist at Humintell (El Cerrito, CA) and Lecturer at Sonoma State University.

Data Availability Statement

The data utilized in this study are available upon reasonable request to the authors for a period of five (5) years after publication, as per guidelines by the American Psychological Association.

References

Addis, D. R., Wong, A. T., & Schacter, D. L. (2007). Remembering the past and imagining the future: Common and distinct neural substrates during event construction and elaboration. *Neuropsychologia, 45*(7), 1363-1377. doi:https://doi.org/10.1016/j.neuropsychologia.2006.10.016

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

Buller, D. B., & Burgoon, J. K. (1996). Interpersonal Deception Theory. *Communication Theory, 6*(3), 203-242. doi:10.1111/j.1468-2885.1996.tb00127.x

Burgoon, J. K., & Buller, D. B. (1994). Interpersonal deception: III. Effects of deceit on perceived communication and nonverbal behavior dynamics. *Journal of Nonverbal Behavior, 18*(2), 155-184. doi:10.1007/bf02170076

Burgoon, J. K., Twitchell, D. P., Jensen, M. L., Meservy, T. O., Adkins, C. T., Kruse, J., . . . Younger, R. E. (2009). Detecting Concealment of Intent in Transportation Screening: A Proof of Concept. *IEEE Transactions on Intelligent Transportation Systems, 10*(1), 103-112. doi:10.1109/TITS.2008.2011700

Cartmill, E. A., & Goldin-Meadow, S. (2016). Gesture. In D. Matsumoto, H. C. Hwang, & M. G. Frank (Eds.), *APA Handbook of Nonverbal Communication* (pp. 307-333). Washington, DC: American Psychological Association.

Christie, R. (1970). Scale construction. In R. Christie & F. L. Geis (Eds.), *Studies in Machiavellianism* (pp. 10-34). New York: Academic Press.

Davis, M., Markus, K. A., Walters, S. B., Vorus, N., & Connors, B. (2005). Behavioral cues to deception vs. topic incriminating potential in criminal confessions. *Law and Human Behavior, 29*(6), 683-704. doi:10.1007/s10979-005-7370-z

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

Dunbar, N. E., Jensen, M. L., Tower, D. C., & Burgoon, J. K. (2014). Synchronization of Nonverbal Behaviors in Detecting Mediated and Non-mediated Deception. *Journal of Nonverbal Behavior, 38*(3), 355-376. doi:10.1007/s10919-014-0179-z

Duran, N., Dale, R., Kello, C., Street, C., & Richardson, D. (2013). Exploring the movement dynamics of deception. *Frontiers in psychology, 4*(140). doi:10.3389/fpsyg.2013.00140

Ekman, P. (1985). *Telling lies: clues to deceit in the marketplace, politics, and marriage* (1st ed.). New York: Norton.

Ekman, P. (2009). *Telling lies: Clues to deceit in the marketplace, politics, and marriage* (4th ed.). New York, NY, US: W. W. Norton & Co, Inc.

Ekman, P., & Friesen, W. V. (1978). *Facial action coding system: Investigator's guide*. Palo Alto, Calif.: Consulting Psychologists Press.

Ekman, P., O'Sullivan, M., Friesen, W. V., & Scherer, K. R. (1991). Invited article: Face, voice, and body in detecting deceit. *Journal of Nonverbal Behavior, 15*(2), 125-135. doi:10.1007/BF00998267

Fenn, E., McGuire, M., Langben, S., & Blandón-Gitlin, I. (2015). A reverse order interview does not aid deception detection regarding intentions. *Frontiers in psychology, 6*(1298). doi:10.3389/fpsyg.2015.01298

Frank, M. G. (2009). Thoughts, feelings, and deception. In B. Harrington (Ed.), *Deception: Methods, Motives, Context and Consequences* (pp. 55-73). Palo Alto, CA: Stanford University Press.

Frank, M. G., & Svetieva, E. (2013). Deception. In D. Matsumoto, H. C. Hwang, & M. G. Frank (Eds.), *Nonverbal Communication: Science and Applications* (pp. 121-144). Thousand Oaks, CA: Sage.

Granhag, P. A., & Mac Giolla, E. (2014). Preventing Future Crimes. *European Psychologist, 19*(3), 195-206. doi:10.1027/1016-9040/a000202

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

Hurley, C. M., & Frank, M. G. (2011). Executing facial control during deception situations. *Journal of Nonverbal Behavior, 35*, 119-131. doi:10.1007/s10919-010-0102-1

Hwang, H. C., & Matsumoto, D. (2016a). Facial expressions. In D. Matsumoto, H. C. Hwang, & M. G. Frank (Eds.), *APA Handbook of Nonverbal Communication* (pp. 257-287). Washington, DC: American Psychological Association.

Hwang, H. C., & Matsumoto, D. (2016b). Measuring emotions in the face. In H. L. Meiselman (Ed.), *Emotion measurement* (pp. 125-144). Cambridge, UK: Elsevier.

Jensen, M. L., Meservy, T. O., Burgoon, J. K., & Nunamaker, J. F. (2010). Automatic, Multimodal Evaluation of Human Interaction. *Group Decision and Negotiation, 19*(4), 367-389. doi:10.1007/s10726-009-9171-0

Johnson, M. K. (1988). Reality monitoring: An experimental phenomenological approach. *Journal of Experimental Psychology: General, 117*(4), 390-394.

Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review, 88*(1), 67-85.

Levine, T. R. (2018). Scientific evidence and cue theories in deception research: Reconciling findings from meta-analyses and primary experiments. *International Journal of Communication, 12*, 2461-2479. doi:1932-8036/201800055

Luke, T. J. (2019). Lessons From Pinocchio: Cues to Deception May Be Highly Exaggerated. *Perspectives on Psychological Science, 14*(4), 646-671. doi:10.1177/1745691619838258

Matsumoto, D., Ekman, P., & Fridlund, A. (1991). Analyzing nonverbal behavior. In P. W. Dowrick (Ed.), *Practical guide to using video in the behavioral sciences* (pp. 153-165). New York: John Wiley & Sons.

Matsumoto, D., & Hwang, H. C. (2015). Differences in word usage by truth tellers and liars in written statements and an investigative interview after a mock crime. *Journal of Investigative Psychology and Offender Profiling, 12*, 199-216. doi:10.1002/jip.1423

Matsumoto, D., & Hwang, H. C. (2018a). Clusters of nonverbal behaviors differ according to type of question and veracity in investigative interviews in a mock crime context. *Journal of Police and Criminal Psychology, 33*(4), 302-315. doi:10.1007/s11896-017-9250-0

Matsumoto, D., & Hwang, H. C. (2018b). Microexpressions differentiate truths from lies about future malicious intent. *Frontiers in psychology, 9*(2545). doi:10.3389/fpsyg.2018.02545

Matsumoto, D., Hwang, H. C., & Frank, M. G. (2016). The body: Postures, gait, proxemics, and haptics. In D. Matsumoto, H. C. Hwang, & M. G. Frank (Eds.), *APA Handbook of Nonverbal Communication* (pp. 387-400). Washington, DC: American Psychological Association.

Matsumoto, D., Hwang, H. C., & Sandoval, V. A. (2015). Ethnic similarities and differences in linguistic indicators of veracity and lying in a moderately high stakes scenario. *Journal of Police and Criminal Psychology, 30*(1), 15-26. doi:10.1007/s11896-013-9137-7

Newman, M. L., Pennebaker, J. W., Berry, D. S., & Richards, J. M. (2003). Lying words: Predicting deception from linguistic styles. *Personality and Social Psychology Bulletin, 29*, 665-675.

Pennebaker, J. W., & Chew, C. H. (1985). Behavioral inhibition and electrodermal activity during deception. *Journal of Personality and Social Psychology, 49*(5), 1427-1433. doi:10.1037/0022-3514.49.5.1427

Schacter, D. L., & Addis, D. R. (2007). The cognitive neuroscience of constructive memory: remembering the past and imagining the future. *Philosophical Transactions of the Royal Society B: Biological Sciences, 362*(1481), 773-786. doi:doi:10.1098/rstb.2007.2087

Schacter, D. L., Addis, D. R., & Buckner, R. L. (2008). Episodic simulation of future events: Concepts, data, and applications. In A. Kingstone & M. B. Miller (Eds.), *Annals of the New York Academy of Sciences* (Vol. 1124: The year in cognitive neuroscience 2008., pp. 39-60). Malden: Blackwell Publishing.

Schacter, D. L., Addis, D. R., Hassabis, D., Martin, V. C., Spreng, R. N., & Szpunar, Karl K. (2012). The Future of Memory: Remembering, Imagining, and the Brain. *Neuron, 76*(4), 677-694. doi:https://doi.org/10.1016/j.neuron.2012.11.001

Scott, S., & McGettigan, C. (2016). The voice: From identity to interactions. In D. Matsumoto, H. C. Hwang, & M. G. Frank (Eds.), *APA Handbook of Nonverbal Communication* (pp. 289-306). Washington, DC: American Psychological Association.

Snyder, M. (1974). Self-monitoring of expressive behavior. *Journal of Personality and Social Psychology, 30*(4), 526-537. doi:10.1037/h0037039

Tsai, J. L., Ying, Y.-W., & Lee, P. A. (2000). The meaning of "being Chinese" and "being American": Variation among Chinese-American young adults. *Journal of Cross-Cultural Psychology, 31*(3), 302-332. doi:10.1177/0022022100031003002

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

Vrij, A., Edward, K., Roberts, K. P., & Bull, R. (2000). Detecting deceit via analysis of verbal and nonverbal behavior. *Journal of Nonverbal Behavior, 24*(4), 239-263. doi:10.1023/A:1006610329284

Vrij, A., Leal, S., Mann, S. A., & Granhag, P. A. (2011). A comparison between lying about intentions and past activities: Verbal cues and detection accuracy. *Applied Cognitive Psychology, 25*, 212-218. doi:10.1002/acp.1665

Wright Whelan, C., Wagstaff, G. F., & Wheatcroft, J. M. (2014). High-Stakes Lies: Verbal and Nonverbal Cues to Deception in Public Appeals for Help with Missing or Murdered Relatives. *Psychiatry, Psychology and Law, 21*(4), 523-537. doi:10.1080/13218719.2013.839931

Zuckerman, M., DePaulo, B. M., & Rosenthal, R. (1981). Verbal and nonverbal communication of deception. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 14, pp. 2-60). New York: Academic Press.

Table 1

Descriptive Statistics (Ms and SDs) and Results of t-tests Comparing Veracity Conditions separately for each NVB

	NVB	Truth N	Truth Mean (SD)	Lie N	Lie Mean (SD)	*t*	df	*p* (2-tailed)	*d*
Face	Anger	99	0.13 (0.44)	113	0.17 (0.50)	-0.57	210	.573	-0.08
	Contempt	99	0.41 (1.53)	113	0.46 (0.97)	-0.27	210	.791	-0.04
	Disgust	99	0.40 (0.82)	113	0.60 (1.41)	1.27	210	.206	-0.18
	Fear	99	0.15 (0.39)	113	0.37 (0.84)	-2.40	210	.017	-0.36
	Happiness	99	0.12 (0.41)	113	0.04 (0.23)	1.91	210	.058	0.27
	Sadness	99	0.15 (0.41)	113	0.24 (0.63)	-1.18	210	.241	-0.17
	Surprise	99	0.12 (0.36)	113	0.05 (0.23)	1.68	210	.095	0.23
Gesture	Head Nods	98	4.29 (4.19)	111	3.95 (4.29)	0.58	207	.564	0.08
	Head Shakes	98	3.68	111	4.36	-1.236	207	.218	-0.17

			(4.01)		(3.90)				
	Shrugs	98	0.59 (1.04)	111	1.23 (1.79)	-3.08	207	.002	-0.45
Whole Body	Body Trembling	62	0.98 (1.58)	73	0.97 (2.71)	0.01	133	.993	0.00
	Body Swaying	62	1.08 (1.33)	73	1.32 (1.78)	-0.85	133	.395	-0.15
	Fidgeting/Grooming	62	1.21 (1.44)	73	1.83 (1.98)	2.10	133	.038	-0.36
	Rigid Posture	62	0.00 (0.00)	73	0.01 (0.06)	-0.92	133	.359	-0.23
Voice	Pitch	96	150.75 (40.19)	111	150.42 (38.89)	0.06	205	.951	0.01
	Pitch Range	96	126.71 (80.96)	111	117.16 (63.72)	0.95	205	.344	0.13
	Intensity	96	62.68 (7.79)	111	64.98 (19.54)	-1.08	205	.281	-0.17
	Intensity Range	96	22.38 (5.17)	111	25.82 (30.74)	-1.08	205	.281	-0.19
	Response Latency	96	0.85 (0.83)	111	0.67 (0.48)	1.92	205	.056	0.27
	Duration and Unfilled Pauses	96	5.95 (3.54)	111	5.03 (3.69)	1.82	205	.070	0.25

Speech and Articulation Rate	96	4.97 (1.84)	111	5.21 (2.03)	-0.89	205	.374	-0.12

Table 2

Results of Binary Logistic Regressions conducted on Facial Expressions of Emotion, Gestures, Whole Body Movements, Voice, and All NVB

NVB Analyzed	Model	df	χ^2	*p*	Classification Accuracy	Predictors	*B*	*SE*	*p*
Facial Expressions of Emotion	3	5	13.83	.017	57.1%	Fear	0.56	0.26	.033
						Surprise	-0.90	0.54	.094
Gestures	2	2	10.32	.006	60.1%	Shrugs	0.36	0.13	.005
Whole Body Movements	1	3	6.93	.074	57.7%	Fidgeting/Grooming	0.29	0.12	.016
Voice	5	4	8.06	.089	60.2%	Response Latency	-0.44	0.26	.090
All NVB	1	21	39.31	.008	76.0%	Disgust	0.567	0.231	.014
						Shrugs	0.432	0.193	.025
						Body Swaying	0.333	0.160	.038
						Fidgeting/Grooming	0.351	0.156	.025
						Pitch Range	-0.011	0.006	.052

						Disgust	0.41	0.20	.038
						Happiness	-1.36	0.77	.077
All NVB	17	5	25.8	< .001	65.9%	Shrugs	0.45	0.17	.007
						Body Swaying	0.24	0.14	.077
						Fidgeting/Grooming	0.36	0.14	.007

Table 3

Post Hoc Tests of Significant Ethnicity Effects

NVB	F	Scheffe comparisons
Facial expressions of anger	$F(3, 180) = 4.64$, $p = .004$, $\eta_p^2 = .072$	CH > HI
Body swaying	$F(3, 107) = 4.80$, $p = .004$, $\eta_p^2 = .119$	CH > EA, HI
Pitch range	$F(3, 175) = 8.85$, $p \leq .001$, $\eta_p^2 = .320$	ME, HI > EA
Intensity	$F(3, 175) = 8.03$, $p \leq .001$, $\eta_p^2 = .121$	ME, HI > CH
Response latency	$F(3, 175) = 3.33$, $p = .021$, $\eta_p^2 = .054$	None
Duration and unfilled pauses	$F(3, 175) = 7.48$, $p \leq .001$, $\eta_p^2 = .114$	CH > EA, HI

Note: CH – Chinese immigrants; EA – European Americans; HI – Hispanic immigrants; ME – Middle Eastern immigrants

Endnotes

[1] Their analyses also showed that within a cluster of NVB, some cues were relatively more important than others and that neither motivation or strong emotion were moderators of the effects. Levine (2018) also commented in his review that arguments about the importance of stakes invoked circular reasoning.

[2] Burgoon et al. (2009) was labeled as an intent study but none of the test beds used in the paper was actually about intent for future malfeasance. It used multiple test beds from experiments about current or past events to make a case for the use of indicators for future intent.

[3] The collection of the original archival video records was approved by the University at Buffalo, State University of New York Social and Behavioral Sciences Institutional Review Board. Written consent was obtained from all interviewees appearing in the video records after they were fully informed about the experimental procedures but before the interviews and once again after the experiment was completed.

[4] Considerably more details than reported below about sampling, recruitment, cultural differences among the immigrant groups, interviewers and questions, stakes, procedures and specific instructions can be found in Matsumoto and Hwang (2015) and Matsumoto et al. (2015).

[5] The criteria allowed us to include individuals who were members of groups that were culturally different (as exemplified by the results on the GEQ below), but who could read, write, and speak English functionally enough to participate in the study. Interested readers are referred to Matsumoto and Hwang (2015) for more details concerning recruitment, discussion of cultural differences, and citations relevant to ethnic differences in nonverbal behavior.

[6] The seven questions were as follows: (1) Good morning/afternoon. What is the purpose of your visit today? (2) Where will you be going? (3) May I see a picture ID? (4) Can you tell me

in as much detail as possible what you plan to do in the file room today? (5) Is that all? (6) Do you intend to engage in any act that involves taking anything that does not belong to you? (7) Is there anything else you wish to tell me about what you plan to do once you pass through this screening? As originally reported in Matsumoto et al. (2015), questions 4, 5, and 6 were those that were diagnostic as truthtellers could answer them truthfully while liars had to lie to them, and which were analyzed in this study.

[7] Sample sizes for specific analyses below differed because of differing missing cases occurring because of technical issues in the various methods of data extractions, differences in source record availability (for whole body movements), or differences in cases with no interview contamination.

[8] Given 21 univariate tests computed, at $\alpha = .05$, 21 x 5% = 1.05 tests should attain an $\alpha \leq .05$ by chance; in reality, four did. At $\alpha = .10$ (trending toward significance), 21 x 10% = 2.10 tests would attain an $\alpha \leq .10$ by chance; in reality, seven did.

[9] But, there are also real differences in depth and quality of memories, as memories of the past have actually occurred and involve the encoding of facts, sensations, emotions, and other associations, little of which exists for future intent. Likely for this reason, cognitive load approaches to detecting lies about the future may not be as effective as lies about the past, as reported by Fenn and colleagues (2015).