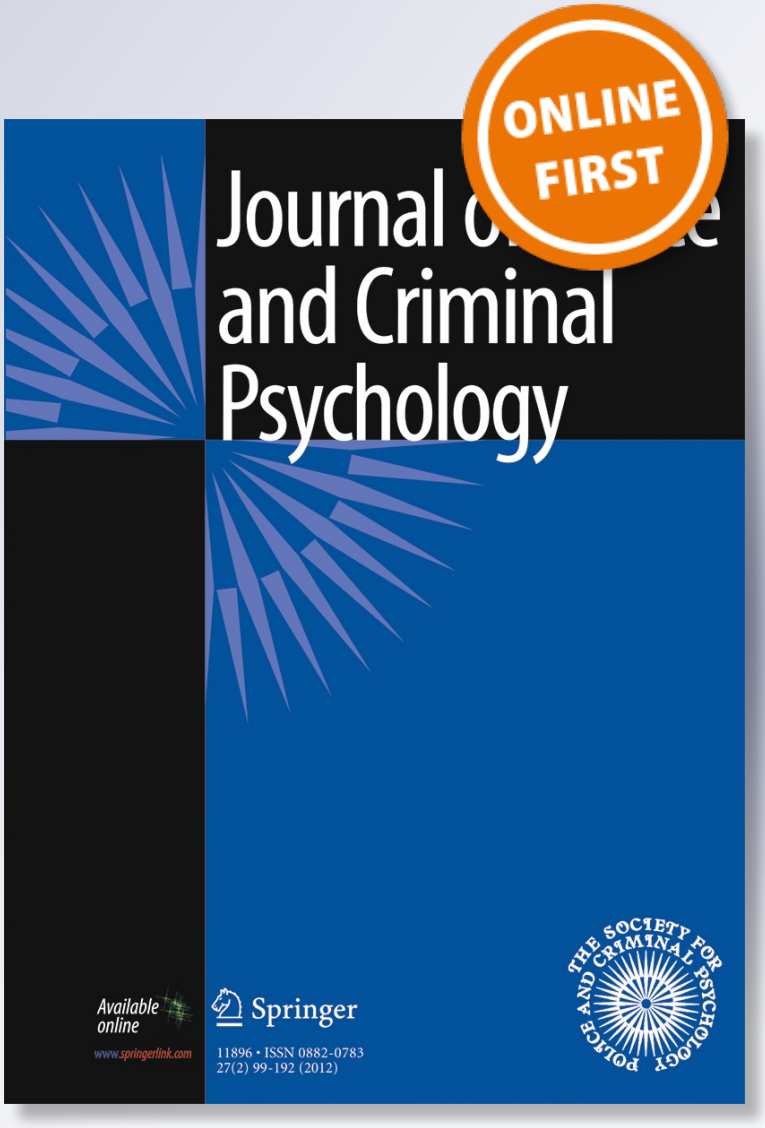
Positive Effects in Detecting Lies from Training to Recognize Behavioral Anomalies

David Matsumoto, Hyisung C. Hwang, Lisa G. Skinner & Mark G. Frank

Journal of Police and Criminal Psychology
The Official Journal of the Society for Police and Criminal Psychology

ISSN 0882-0783

J Police Crim Psych
DOI 10.1007/s11896-012-9115-5



Springer

J Police Crim Psych
DOI 10.1007/s11896-012-9115-5

Positive Effects in Detecting Lies from Training to Recognize Behavioral Anomalies

David Matsumoto · Hyisung C. Hwang · Lisa G. Skinner · Mark G. Frank

Abstract We examined whether training in both the verbal and nonverbal indicators of truth telling and lying would have positive effects on Law Enforcement Officers' (LEOs) ability to evaluate truths from lies. College course-level training on empirically validated verbal and nonverbal indicators of truth telling and lying was provided to mid- to advanced-career level LEOs, whose accuracy in detecting lies from truths was assessed pre- and post-training using truthful and deceptive videos of mock crimes and opinions. A marginally significant truth bias existed at pre-test; training, however, resulted in a significant improvement in accuracy rates for both truth and lie videos, and the truth bias that existed at pre-test was eliminated. Additional analyses indicated that accuracy rates improved for videos of mock crimes but not for opinions. These findings add to a small but growing literature that indicates that training on validated verbal and nonverbal indicators of truth telling and lying has positive benefits.

Keywords Lie detection · Training · Statement analysis · Nonverbal behaviors · Microfacial expressions

Portions of this report were prepared with the support of research grants FA9550-09-1-0281 and FA9550-11-1-0306 from the Air Force Office of Scientific Research.

D. Matsumoto (✉) · H. C. Hwang
Department of Psychology, San Francisco State University and Humintell, LLC, 1600 Holloway Avenue,
San Francisco, CA 94132, USA
e-mail: dm@sfsu.edu

L. G. Skinner
Federal Bureau of Investigation, Washington, DC, USA

M. G. Frank
Department of Communication, University at Buffalo, State University of New York, Buffalo, NY, USA

Published online: 17 November 2012

Identifying indicators associated with truth telling and lying can be a valuable investigative aid for law enforcement officers (LEOs), and research has provided rich sources of information concerning such indicators. One comes from the analysis of the words used when providing statements or answering questions. This body of evidence centers on a technique known as Statement Analysis (SA), which has its roots in the work of Undeutsch (1989; also known as Statement Validity Analysis, Criteria Based Content Analysis, Reality Monitoring, Scientific Content Analysis, and others). SA is based on principles concerning the nature of human memory and its recall and suggests that lies are different from truths in their demands upon memory, which are reflected by changes in the use of some aspects of grammar and language. Lies contain fewer words and omissions of information; are less plausible, structured, and logical; are more internally discrepant and ambivalent; contain repeated details and lack contextual embedding; and include more descriptions of what did not occur (DePaulo et al. 2003; Duran et al. in press; Newman et al. 2003; Porter et al. 2000a; Porter and ten Brinke 2010; Vrij 2007).

Another source of indicators comes from a body of work examining nonverbal behaviors (NVB). NVB indicators occur because emotions and cognitions are expressed nonverbally and because conflicting thoughts and feelings that occur when lying often leak out despite attempts to control them. NVB indicators of lying include changes in the use of speech illustrators and emblematic gestures; facial expressions of emotion and cognition, especially emotional leakage; changes in blinking, pauses, and speech rates; and attempts to control or regulate one's emotions (DePaulo et al. 2003; Hurley and Frank 2011; Porter and ten Brinke 2008, 2010; Porter et al. 2012; ten Brinke et al. 2012).

Despite the existence of these two rich sources of indicators of truth telling and lying, they are not readily recognizable. People are generally poor at discerning lies from

truths, with average accuracy rates around 54 % (C. F. Bond and DePaulo 2006, 2008); studies with LEOs have generated similar results (Vrij and Mann 2001). This makes sense because in everyday life society requires a certain degree of blindness to others' true thoughts and feelings in order to function effectively (Matsumoto et al. 2008). But for some professions it is crucial that lies are detected more accurately than chance; law enforcement is one such profession.

Given that clues to deception do exist, one important question concerns whether individuals in general, and LEOs in particular, can be trained to improve their ability to detect lies. A meta-analysis of 11 training studies (two involved LEOs) reported a modest positive impact of training, despite limitations in the training techniques found across the studies; e.g., a majority of the studies trained for under one hour, or trained on unsubstantiated behaviors (Frank and Feeley 2003). A subsequent review examined nine additional studies of 16 trained groups (three involved LEOs) and reported much more substantial gains in lie detection accuracy often using more precise techniques (O'Sullivan et al. 2010). Other studies not included in these two reviews also showed positive benefits of training of indicators to lie detection accuracy (Landry and Brigham 1992; Porter et al. 2010; Porter et al. 2000b).

But despite the fact that two major sources of indicators exist, training programs typically focus only on one or the other. In reality both co-occur simultaneously, especially in face-to-face interactions but even when interviewees are alone writing statements. Thus it is important to train individuals to recognize verbal and nonverbal indicators of truth telling and lying in combination with each other because neither occurs in a vacuum and both always co-occur. Training in both sources of indicators is therefore more ecologically valid because it encourages trainees to attend to both the words and NVB, which is more analogous to real-life settings. Training in both sources also considerably expands the toolkits available to LEOs when conducting interviews and interrogations. A recent study demonstrated that combining both verbal and nonverbal indicators of deception resulted in an additive effect of predicting truths from lies (Matsumoto et al. 2011).

We examined whether training in both the verbal and nonverbal indicators of truth telling and lying would have positive effects on LEOs' ability to evaluate truths from lies. This study contributes to a small literature examining the effect of training of both verbal and nonverbal clues to deception (Porter et al. 2010; Porter et al. 2000b), and is unique in its inclusion of professional LEOs as trainees. The training delivered addressed guidelines for research examining the efficacy of lie detection training (Docan-Morgan 2007; Frank and Feeley 2003; O'Sullivan et al. 2010), foremost of which concerns the training content – that trainees should be taught about indicators empirically validated to be indicative of truths or lies. Indeed, being taught non-validated indicators could have dire consequences, especially if the detection of deception is associated with guilt. For example, many people believe that a lack of eye contact is an indicator of lying, and this belief is held across cultures (The Global Deception Research Team 2006). But a number of studies have tested this possibility and most do not support it (Frank and Svetieva 2013); thus this belief is more myth than reality. It is no wonder, then, that training programs teaching invalid indicators produce negative results (e.g., Kassin and Fong 1999). Similar non-findings for techniques based on non-validated SA techniques have also been obtained (Smith 2001).

The training provided here covered empirically validated verbal and nonverbal indicators of truth telling and lying and was delivered in a professional setting within an advanced training curriculum for mid- to advanced-career level LEOs. The videos testing pre- and post-training lie detection accuracy were relevant to LEOs and involved high-stakes situations with emotionally-laden consequences for being believed or not. These characteristics ensured that the training content was accurate, empirically sound, and relevant for LEOs, and the study was the first to combine both verbal and nonverbal indicators in a college course-level training of professional law enforcement officers using scientifically backed clues to deceit. Trainee's accuracies in detecting lies from truths were assessed pre- and post-training. We hypothesized that truth and lie accuracy scores would improve significantly from pre- to post- as a result of training.

Methods

Participants

Participants were all trainees of the Federal Bureau of Investigation's (FBI) National Academy (NA) located in Quantico, VA. The FBI NA is a professional course of study for U.S. and international law enforcement leaders that serves to improve the administration of justice in police departments and agencies in the U.S. and elsewhere, and to raise law enforcement standards, knowledge, and cooperation worldwide. Leaders and managers of state and local police, sheriffs' departments, military police organizations, and federal law enforcement agencies attend by invitation though a nomination process. Participants are drawn from the U.S., U.S. territories, and over 150 international partner nations.

Participants were concurrently enrolled in two courses that occurred in three separate NA sessions (Ns=16, 18, and 22 for sessions 1, 2, and 3, respectively; total N across all three sessions=56). The courses were entitled "Analysis of Verbal and Nonverbal Behavior Communication" (AVNBC) and "Interviewing Strategies through Statement Analysis"

(ISSA). The samples were predominantly male (N_{male}=52) and from U.S. law enforcement agencies ($N_{U.S.}$ = 45). They were all senior personnel with an average 17.7 years of experience as a LEO.

Course Descriptions

Each course occurred over a 10-week period that allowed for 22 two-hour class sessions. The ISSA involved lectures with discussion and group projects. Lectures included specific concepts and linguistic features empirically demonstrated to differentiate truths from lies (Adams and Harpster 2008; Adams and Jarvis 2006; Johnson 1988; Johnson and Raye 1981; Masip et al. 2005; Newman et al. 2003; Porter and Yuille 1996; Ruby and Brigham 1997; Vrij 2007; Vrij and Mann 2006; Zaparniuk et al. 1995) and covered the following topics: identifying the incident; calculating balance, elements of time, equivocation, negation, and extraneous information; unique sensory detail; emotion; persons in order of appearance; and the use of nouns, verbs, adverbs, and adjectives. The group projects involved analysis of statements obtained during law enforcement investigations from victims and witnesses of violent crimes, and suspects accused of violent or property crimes. Trainees were instructed to analyze each statement using the previously listed concepts and linguistic features in order to gain insight into the author's thoughts, motivations, and ideas.

The AVNBC course involved lectures with discussion, web-based exercises, video review exercises, and practica with role players and each other. The specific nonverbal behaviors covered were those empirically associated with deception and included facial expressions of emotion and cognition (Ekman et al. 1991; Frank and Ekman 1997; Hurley and Frank 2011; Porter and ten Brinke 2008; ten Brinke et al. 2012; Warren et al. 2009); gestures (deTurck and Miller 1985; Ekman and Friesen 1969); vocal cues including speech errors, pauses, pitch changes, response latencies, and speech durations (DePaulo et al. 2003; deTurck and Miller 1985; Kraut 1978; Zuckerman et al. 1981); and changes in adaptors (Ekman and Friesen 1969; Zuckerman et al. 1981). Other lecture topics included an overview of evaluating truthfulness and detecting deception, active listening and psychological persuasion, and interview and interrogation considerations. The web-based exercises included a basic online training course in how to recognize microfacial expressions of emotion (MiX; see Matsumoto and Hwang 2011), a more advanced course that showed faces at three different angles (MiX Elite; Ekman and Matsumoto 2008), and a course in how to recognize subtle facial expressions of emotion (SuBX Professional; Matsumoto and Hwang 2010). These training tools presented universal facial expressions of emotion for brief durations (100 msec), with a forward and backward mask of the same expressor's neutral face. The training tools all included a pre-test, an instructional section in which each of the expressions are introduced and described, a practice section, a review section, and a post-test. Trainees achieved a minimum of 80 % accuracy on the post-tests of each of these courses. Because the goal of this course was to combine both SA and NVB techniques, exercises included reviews and discussions of verbal and nonverbal behavior shown in multiple videos highlighting the differences between truth-tellers and liars from known ground-truth examples, as well as reviews and discussions of videos of actual police interviews and interrogations. The practica included a truth-lie exercise that trainees did at the beginning of the semester in which they either told the truth or a lie about a personal experience while being videotaped, which they watched and analyzed later in the semester. There was also a practicum exercise at the end of the semester that involved the trainee's interview and interrogations of contracted role players.

Lie Detection Videos

The videos that comprised the lie detection tests were the same used in previous studies (Ekman et al. 1999; Frank and Ekman 1997) and came from two sets of video pools, both created to represent ecologically-valid, high-stake scenarios. One was an adaptation of a *mock crime scenario* in which individuals were instructed that they and a second person (a confederate) would enter a room, one at a time, and have the opportunity to steal $50 cash from inside a briefcase. The person who entered first could choose whether or not to take the money; the person who entered second would have to take the money if it was there. The order of the individual entering the room was counterbalanced. All individuals knew that they would be interrogated concerning the theft and were instructed to deny that they had taken the money. If they lied but were believed by the interrogator, they could keep the $50 as a bonus above their participation fee ($10); if they lied but were not believed, they would forfeit not only the bonus money but also their participation fee. Half the individuals were also told they would have to sit on a cold, metal chair inside a cramped, darkened room labeled *XXX,* where they would have to endure 10 to 40 randomly sequenced, 110-decibel blasts of white noise over the course of 1 hour; they were given a sample of this punishment prior to engaging in the task but no participant actually received the punishment. If the individual told the truth and was believed, he would receive a $10 bonus on top of his participation fee. If the individual told the truth but was not believed, he would face the same punishment as the participant judged as lying.

The second scenario was the *opinion scenario*, which was based on the false opinion paradigm (Mehrabian 1971). Participants were given a questionnaire that assessed

both the direction and magnitude of their opinions on social issues (e.g., "Should convicted coldblooded murderers be executed?" and "Should smoking be banned from all enclosed public places?") on a 1 *(strongly agree)* to 7 *(definitely disagree)* rating scale. The opinion that the participant felt most strongly about—and hence was most certain about—was selected as the one to be described truthfully or falsely. A false description meant trying to convince the interrogator that the participant held an opinion on that issue exactly opposite to the participant's true opinion. The stakes for this scenario were identical to those of the crime scenario.

Video recordings of 20 persons – 10 for the crime scenario and 10 for opinion – were edited to show their responses to the last five (for Opinion) or six (for Crime) questions asked.[1] No person appeared on more than one video and each scenario included an equal number of liars and truth-tellers. Each video clip was approximately 60 – 90 s in length and showed a facial close-up with full audio. The interrogator also could be heard—but not seen—asking his questions.

The videos were shown to both student and non-student observers in multiple previous studies (Ekman, et al. 1999; Frank and Ekman 1997) who were asked to judge whether the individual was telling the truth or lying. Because ground truth for each person was known, norm accuracy scores based on naïve undergraduate observers were generated for each clip. We produced two tests (Tests A and B) that were used in this study (10 video clips each); both included an equal number of mock crime and opinion videos and an equal number of liars and truth-tellers. More importantly both tests had the same mean norm accuracy score (.56 for Test A, .58 for Test B); these are comparable to norms lie detection accuracy found in previous research (C. F. Bond and DePaulo 2006).

Procedures

Tests A and B served as either pre- or post-tests. Test A was the pre-test for the first and third sessions while Test B was the pre-test for the second. The matching of the norm accuracy means for Tests A and B and the varying of the order of the tests ensured that differences in pre-post scores could not be attributed to differences in the inherent difficulty of the tests.

On the first day of the course trainees were provided a URL that directly linked them to the pre-test, which they were to complete during their free time on a high-speed Internet connection before the start of the first class the second week. Trainees completed the test online and independently of the course, and the course instructors did not know the scores for individual trainees in the class. The course was then administered as planned over the subsequent weeks. During the last week of the course, trainees were again provided a URL that directly linked them to the post-test, which the trainees were able to complete until the last day of class. All trainees completed all tests individually and independently of each other, not in groups with any discussion. Scores on the post-test were calculated independently of the grades for the course.

After answering some basic demographic questions, trainees were provided with the following standard instructions when doing the tests:

> "You will see 10 videos of brief interviews with individuals who are either lying or telling the truth. The first five videos will be interviews with individuals who are either telling the truth or lying about whether they committed a CRIME (stole some money or not). The second five videos will be interviews with individuals who are either telling the truth or lying about their OPINIONS about something (smoking bans, death penalty, abortion, etc.). Your job is to evaluate whether the individual is telling the truth or lying, and to describe the reasons why you came to your decision. Describe any and all reasons why you came to that decision, including your evaluations of the individual's nonverbal behaviors or statements."

Trainees clicked "Next" when they were ready to view a video. After a video played, they were asked "Was this person telling the truth or lying? Please select" and made a dichotomous Truth or Lie judgment. We recoded each judgment into an accuracy score and computed the percentage of accurate judgments for each trainee for each test; thus scores could range from 0 to 100. Because of technical problems, one video clip from Test B for the second session was not used; thus scores for that group's test were computed using 9 clips.

Results

Overall Analyses

We computed a two-way, repeated measures ANOVA on the overall accuracy scores across all video items using Time (pre- vs. post-test) and Session (3) as independent variables. The main effect of Time was significant, $F(1, 51)=16.20$, $p<.001$, $\eta_p^2=.24$, indicating that the group as a whole increased in their

[1] For the Crime videos the questions were: (1) Can you describe exactly what happened, what you saw and did, when you went in that room? (2) Describe what your thoughts were when you stood in that room and looked in that envelope. (3) Do you know how much money was or was supposed to be in the envelope? (4) Did you take that money from the envelope? (5) Did you bring the money with you into this room? (6) Are you lying to me right now? For the Opinion videos the questions were: (1) What is your opinion on this issue? (2) Why do you believe that? (3) How long have you had this opinion? (4) Did you just make it up a few minutes ago? (5) Are you lying to me now?

accuracy scores from pre- (M=51.30, SD=13.10) to post-test (M=61.85, SD=15.18). The 10 % gain in accuracy is substantial, given that the video clips were only 60 to 90 s long. The effect size of .24 corresponds to an r of .49, and indicated that a substantial proportion of the variance in accuracy scores was accounted for by improvement from pre- to post. The pre-test mean indicated that the samples were not different than the norm accuracy data for the test videos, eliminating the possibility of sample bias affecting the positive results.

The Time by Session interaction was not significant, $F(2, 51)=.12$, $p=.89$, $\eta_p^2=.01$, indicating that the improvement did not vary across sessions. Regardless, we tested the differences between the pre- and post-test scores separately for each of the three sessions. Because of the small sample sizes we utilized sign tests and Wilcoxin signed-rank tests. Post-test scores were significantly higher than pre-tests for all three sessions, all ps<.05 for both tests.

Truth vs. Lie Videos

We computed separate accuracy scores for Truth and Lie videos and examined the differences between them as a function of training (Table 1). Because previous studies have documented a truth bias in judgments of lie situations, especially when using short video clips (Feeley et al. 1995; Levine et al. 1999; Masip et al. 2009b), we tested the difference between the Truth and Lie videos at the pre-test. There was a marginally significant difference between accuracy rates for Truth and Lie videos, $t(53)=1.95$, $p<.06$, $d=.26$, with accuracy rates for Truth videos being slightly higher than that for Lie videos, supporting the idea of a possible truth bias at the beginning of the training. Simple effects analyses of the training effect (pre- vs. post-test), however, indicated that the accuracy rates for *both* increased as a result of training, $t(53)=1.84$, $p<.05$, $d=.25$; and $t(53)=4.31$, $p<.001$, $d=.59$, for Truth and Lie videos, respectively. Moreover, the difference between the Truth and Lie videos was not significant at post-test $t(56)=.59$, ns. Inclusion of Session in a three-way ANOVA (with Time and Video Type as the other factors) did not produce any significant effects involving Session. Thus, not only were the training effects replicated for both Truth and Lie videos; they also indicated a reduction in any pre-session bias the participants had in their ability to discern truths from lies. Also, the training effects were larger for Lie than Truth videos.

Table 1 Descriptive Statistics for Accuracy Scores separately for Truth and Lie Videos

	Truth Videos	Lie Videos
Pre-Test	54.81 (18.22)	47.78 (19.20)
Post-Test	61.11 (21.07)	62.59 (19.44)

Top value in each cell is the mean; bottom value in parenthesis is the corresponding standard deviation

Crime vs. Opinion Videos

We also computed separate accuracy scores for the Crime and Opinion videos. At pre-test, accuracy scores for Crime videos (M=42.96, SD=21.42) were significantly *lower* than those for Opinion videos, (M=60.09, SD=18.82), $t(53)=4.12$, $p<.001$, $d=.56$. Simple effects analyses (pre- vs. post-test) indicated that the accuracy rates for Crime videos increased as a result of training, ($M_{post\text{-test}}=68.07$, SD=21.00), $t(53)=6.10$, $p<.001$, $d=.83$; but there was no significant change for Opinion videos, ($M_{post\text{-test}}=54.04$, SD=21.37), $t(53)=1.37$, ns. Moreover, at post-test participants had significantly higher scores on Crime videos than they did on Opinion videos, $t(56)=3.65$, $p<.001$, $d=.48$. Thus the training effects may have been localized to the Crime videos.

Post-Hoc Analyses

Some studies in the literature have suggested that just increasing the amount of active processing of verbal and nonverbal behaviors can lead to improvements in accuracy judgments regardless of having been trained to identify valid indicators of truth and lies (Bond et al. 2004; Levine, et al. 1999). And at least one study has reported that a group that received placebo training on invalid behavioral indicators did just as well as a group that received training on valid indicators, presumably because of increases in active processing of the stimuli (Levine et al. 2005). To test whether this may have accounted for increases in the accuracy rates observed in this study, we examined the amount of time trainees spent completing the pre- and post-tasks, considering the time spent as a proxy for active processing. Across the entire sample the time spent by the trainees did not differ between the pre- and post-tests, $t(56)=1.08$, ns. These data, however, included individuals whose times were inordinately long, probably because they neglected to log out of the survey. But even when the data were filtered to include only those individuals who spent less than two hours on either test, there were still no differences between the pre- (M=52.84 minutes, SD=29.42) and post-tests (M=59.49 minutes, SD=25.18), $t(41)=1.08$, ns. These data suggested that the improvements in accuracy rates did not occur because of increases in active processing of the stimuli.

Discussion

The analyses generated several findings of note. When the videos were analyzed as a collective whole, there was an overall increase in accuracy rates of about 10 %, which corresponded with a fairly large effect size. The positive

training effects were replicated when data were analyzed separately for Truth vs. Lie videos, and a modest pre-session truth bias exhibited at pre-test was eliminated at post-test. Interestingly the training effect was larger for Lie videos than Truth videos, and localized to Crime videos as opposed to Opinion videos. The training effect size for Crime videos, in fact, was substantially large, with a 25 % increase in accuracy rates. These latter findings speak to the effectiveness of the training to the most relevant types of lies LEOs deal with.

This study was not conducted without limitations, perhaps the largest of which was the lack of a control group. Given that the training occurred within classes offered at the FBI NA, it was logistically impossible to include a wait-list or placebo control or comparison group. Because of this we cannot entirely rule out the possibility that trainees improved in their lie detection abilities simply by attending a course at the FBI NA, nor can we infer that the effects were exclusively dependent on this particular course content. However, there is no reason that any other course not specifically designed to improve the recognition of indicators of veracity or lying would produce comparable improvements or that a placebo training would produce improvements, especially given the fact that all trainees were senior personnel who typically have attended many training workshops on interviewing and interrogation, and in some cases training involving SA or NVB.

Other limitations included the relatively small number of videos used in the pre- and post-tests and their brevity, and the fact that the trainees were mere observers of the interaction and did not have the freedom to question the interviewees themselves. Because of these limitations, we consider the task a fairly impoverished one, in which trainees had to make judgments based on minimal information available to them. Despite that major limitation, we interpret the 10 % improvement in overall accuracy rates (and 25 % for Crime videos) as quite remarkable. Although accuracy rates of 60-65 % in this task appear low, we believe that if the trainees had the freedom to question the interviewees themselves and had many of the other sources of evidence typically available to them such as forensics, witness statements, physical evidence, and so forth, the improvements obtained could well translate to fairly substantial differences in the efficacy by which ground truth can be obtained and cases closed. Moreover, 10 pre-test and 10 post-test items, or even fewer, have been used in other training studies in the past. For example, Frank and Feeley (2003) reported that 8 of the 11 studies in their meta-analysis of training studies published before the year 2000 contained less than 10 stimulus items in the pre or post tests. Thus the number of stimulus items is not extraordinarily small by the standards of the published literature.

Another potential limitation of the videos used to assess lie detection accuracy was the fact that they portrayed undergrad and graduate students and some members of the community, while the videos used in the training included actual victims, suspects, and witnesses. We do not believe, however, this characteristic of the videos to be a limiting factor because there is no evidence to suggest that verbal and nonverbal leakage is age specific in adulthood, and there is no reason or evidence to suggest that the behavioral anomalies portrayed in the videos were different than those produced by the individuals used in the videos for training in the classes. And in any case the age differences between the individuals in the test videos (early to late-20s) was not that different than many of the young adults in the training videos. The use of undergrads in videos for testing lie detection accuracy is common (e.g., Porter et al. 2000b) and the same videos used in this study have been used to test lie detection accuracy in LEOs previously (Ekman et al. 1999; O'Sullivan and Ekman 2004).

That improvements in lie detection accuracy occurred when SA and NVB techniques were combined may appear to be common sense given that training on either approach generally produces positive benefits. We do not believe, however, that just because each technique independently produces positive effects they should necessarily do so in combination. Trainees learning both techniques independently often report that they are overwhelmed by the amount of detail to which they need to pay attention, and it was very possible that training in both could have produced too much of an information overload such that their practical application may not have been successful. This possibility was certainly reflected, albeit anecdotally, in remarks made by many of the trainees, especially earlier in the sessions. Trainees reported that they were much more comfortable with applying those techniques by the end of the sessions, but we believe that incorporating the techniques into their behavioral routines heuristically would still require some time after the sessions ended. Future studies should follow trainees post-training to document the retention rates of the positive effects of training.

Analyses of the amount of time spent on the pre- and post-tests suggested that the improvements in accuracy rates were not merely a function of increases in active processing of the video stimuli. Moreover a previous study examining microexpression recognition training using the exact same training tool used in this study (MiX) also demonstrated that individuals trained to see microexpressions significantly improved in their ability to recognize them, even compared to a group that received placebo training (akin to the "bogus control group" used in Levine, et al. 2005), and retained their increased abilities even when tested several weeks later (Matsumoto and Hwang 2011). Our non-findings on time spent, coupled with the previous documented evidence for the positive effects of microexpression training, suggested that the improvements rates observed in this study were not merely due to increases in active processing of the stimuli.

The findings did support the existence of a modest truth bias in the trainees at pre-test, replicating previous similar findings (Feeley et al. 1995; Levine et al. 1999; Masip et al. 2009b). It is interesting that the truth bias existed in the samples here, given our anecdotal observation (and the suggestion of one of the reviewers) that LEOs exhibit a lie, not truth, bias. At least one study has reported that, because of the truth bias, training on lie detection may impact accuracy for lies but not truths (Masip et al. 2009a). Our findings, however, did not support this idea, as the training produced an increase in accuracy rates for both Truth and Lie videos (although the effect sizes were larger for the Lie videos, but this might have been because of the higher accuracy rates for Truth videos at pre-test).

Unfortunately the design of the study precluded us from knowing exactly what the crucial ingredients of the courses were that produced a positive benefit in lie detection accuracy. Although the lengths of the courses were suitable for settings such as the FBI NA or other academic institutions, they are not scalable to larger communities of LEOs. Future research needs not only to replicate the current findings but also to identify the minimum combination of elements in SA and NVB training that can produce a positive benefit for lie detection to produce shorter, more focused training that is scalable to a large workforce.

Recognizing behavioral anomalies in verbal and NVB can not only aid investigators in detecting lies more accurately; they can also be used as aids during interviews and interrogations to help the investigator gain insights about the personality, motivation, and internal conflicts of their interviewees, and to identify meaningful content areas of the interview that deserve further exploration and discovery. These techniques need to be imbedded within a thoughtful and strategic interview and interrogation methodology that has at its ultimate goal the uncovering of ground truth.[2] Recognizing the behavioral anomalies in verbal statements and NVBs are signs that investigators can use to help them get to ground truth; but those signs should not be interpreted as signals of ground truth as is, given that no research has identified any behavior or behavior combination unique to deception (Zuckerman, et al. 1981). Thus they are better deployed as means to an end rather than an end in and of themselves (Frank et al. 2006) and are highly dependent upon the interviewing skills of the investigator.

Using behavioral anomalies to evaluate truthfulness and detect lies in investigative interviewing is not a silver bullet that will solve every case. As always interviews and interrogations need to be augmented by other sources of evidence such as witness statements, forensics, and other evidence. Investigators must still prepare and plan for interviews and interrogations, and craft questions and guide discussions when anomalies are recognized. Additionally investigators who have received training in lie detection need to be cautious of post-training biases (Masip et al. 2005). Yet, identifying valid behavioral anomalies that indicate truth telling and lying – both verbally and nonverbally – can be an incredibly useful aid for any investigator.

[2] The approach that we advocate for the use of behavioral anomalies is compatible with the PEACE model of investigative interviewing. PEACE is an acronym for Preparation and Planning, Engage and Explain, Account, Closure, and Evaluation, and is a set of guidelines for non-confrontational investigative interviewing proposed by the Royal Commission on Criminal Justice in the U.K. in 1993. The PEACE approach minimizes accusatory interrogation tactics that are often adopted if individuals are suspected of lying, as inferences about lying are often associated with inferences about guilt. To be sure, however, the correct interpretation of behavioral anomalies can be used in any investigative interviewing procedure, as they are ultimately used not to draw conclusions about veracity, but instead to draw attention to topics to pursue through more questions (Frank et al. 2006).

References

Adams SH, Harpster T (2008) 911 Homocide calls and statement analysis. Fed Bureau of Investig Law Enforcement Bull 77 (6):22–31

Adams SH, Jarvis JP (2006) Indicators of veracity and deception: an analysis of written statements made to police. Speech Lang and the Law 13(1):1–22

Bond CF, DePaulo BM (2006) Accuracy of deception judgments. Personal Soc Psychol Rev 10:214–234

Bond CF, DePaulo BM (2008) Individual differences in judging deception: accuracy and bias. Psychol Bull 134(4):477–492

Bond GD, Malloy DN, Thompson LA, Arias EA, Nunn SN (2004) Post-probe decision making in a prison context. Commun Monogr 71(3):269–285

DePaulo BM, Lindsay JJ, Malone BE, Muhlenbruck L, Chalrton K, Cooper H (2003) Cues to deception. Psychol Bull 129(1):74–118

deTurck MA, Miller GR (1985) Deception and arousal: isolating the behavioral correlates of deception. Hum Commun Res 16:603–620

Docan-Morgan T (2007) Training law enforcement officers to detect deception: a critique of previous research and framework for the future. Appl Psychol in Crim Justice 3(2):143–171

Duran ND, McCarthy PM, Hall C, McNamara DS (2010) The linguistic correlates of conversational deception: Comparing natural language processing technologies. *Applied Psycholinguistics* 31:439–462

Ekman P, Friesen WV (1969) Nonverbal leakage and clues to deception. Psychiatry 32:88–105

Ekman P, Matsumoto D (2008) Microexpression Recognition Training Tool, Elite (MiX Elite). Available at http://www.humintell.com

Ekman P, O'Sullivan M, Friesen WV, Scherer KR (1991) Invited article: face, voice, and body in detecting deceit. J Nonverbal Behav 15(2):125–135

Ekman P, O'Sullivan M, Frank MG (1999) A few can catch a liar. Psychol Sci 10(3):263–266

Feeley TH, deTurck MA, Young MJ (1995) Baseline familiarity in lie detection. Commun Res Rep 12:160–169

Frank MG, Ekman P (1997) The ability to detect deceit generalizes across different types of high-stake lies. J Personal Soc Psychol 72:1429–1439

Frank MG, Feeley TH (2003) To catch a liar: challenges for research in lie detection training. J Appl Commun Res 31(1):58–75

Frank MG, Svetieva E (2013) Deception. In: Matsumoto D, Frank MG, Hwang HS (eds) Nonverbal communication: science and applications. Sage, Newbury Park, pp 121–144

Frank MG, Yarbrough JD, Ekman P (2006) Improving interpersonal evaluations: combining science and practical experience. In: Williamson T (ed) Investigative interviewing: rights, research, regulation. Willan Publishing, Portland, pp 229–255

Hurley CM, Frank MG (2011) Executing facial control during deception situations. J Nonverbal Behav 35:119–131

Johnson MK (1988) Reality monitoring: an experimental phenomenological approach. J Exp Psychol Gen 117(4):390–394

Johnson MK, Raye CL (1981) Reality monitoring. Psychol Rev 88 (1):67–85

Kassin SM, Fong CT (1999) "I'm Innocent!": effects of training on judgments of truth and deception in the interrogation room. Law Hum Behav 23:499–516

Kraut RE (1978) Verbal and nonverbal cues in the perception of lying. J Personal Soc Psychol 36(4):380–391

Landry KL, Brigham JC (1992) The effect of training in criteria-based content analysis on the ability to detect deception in aduts. Law Hum Behav 16(6):663–676

Levine TR, Park HS, McCornack SA (1999) Accuracy in detecting truths and lies: documenting the "veracity effect". Commun Monogr 66:125–144

Levine TR, Feeley TH, McCornack SA, Hughes M, Harms CM (2005) Testing the effects if nonverbal behavior training on accuracy in deception detection with the inclusion of a bogus training control group. West J Commun 69(3):203–217

Masip J, Sporer SL, Garrido E, Herrero C (2005) The detection of deception with the reality monitoring approach: a review of the empirical evidence. Psychol Crime and Law 11(1):99–122

Masip J, Alonso H, Garrido E, Herrero C (2009a) Training to detect what? the biasing effects of training on veracity judgments. Appl Cogn Psychol 23:1282–1296

Masip J, Garrido E, Herrero C (2009b) Heuristic versus systematic processing of information in detecting deception: questioning the truth bias. Psychol Rep 105(1):11–36

Matsumoto D, Hwang HS (2010) Subtle expression recognition training tool: SuBX Professional. Available at http://www.humintell.com

Matsumoto D, Hwang HS (2011) Evidence for training the ability to read microexpressions of emotion. Motiv Emot 35(2):181–191

Matsumoto D, Yoo SH, Fontaine JRJ, Anguas-Wong AM, Arriola M, Ataca B, Grossi E (2008) Mapping expressive differences around the world: the relationship between emotional display rules and individualism v. Collectivism. J Cross-Cult Psychol 39(1):55–74

Matsumoto D, Hwang HS, Skinner L, Frank MG (2011) Evaluating truthfulness and detecting deception. *FBI Law Enforcement Bulletin, June*, 1–11

Mehrabian A (1971) Nonverbal betrayal of feeling. J Exp Res Personal 5:64–73

Newman ML, Pennebaker JW, Berry DS, Richards JM (2003) Lying words: predicting deception from linguistic styles. Personal Soc Psychol Bull 29:665–675

O'Sullivan M, Ekman P (2004) The wizards of deception detection. In: Granhag PA, Stromwall LA (eds) The detection of deception in forensic contexts. Cambridge University Press, London, pp 269–286

O'Sullivan M, Frank MG, Hurley CM (2010) Police lie detection: accuracy, training, and new directions. In: Peters JM (ed) Police psychology. Nova Science Publishers, Hauppauge, pp 75–88

Porter S, ten Brinke L (2008) Reading between the lies: identifying concealed and falsified emotions in universal facial expressions. Psychol Sci 19(5):508–514

Porter S, ten Brinke L (2010) The truth about lies: what words in detecting high-stakes deception? Leg Criminol Psychol 15:57–75

Porter S, Yuille JC (1996) The language of deceipt: an investigation of the verbal clues to deception in the interrogation context. Law Hum Behav 20:443–459

Porter S, Birt AR, Yuille JC, Lehman DR (2000a) The negotiation of false memories: the influence of interviewer and rememberer characteristics on memory distortion. Psychol Sci 11:513–516

Porter S, Woodworth M, Birt AR (2000b) Truth, lies, and videotape: an investigation of the ability of federal parole officers to detect deception. Law Hum Behav 24(6):643–658

Porter S, Juodis M, ten Brinke L, Klein R, Wilson K (2010) Evaluation of the effectiveness of a brief deception detection training program. The J of Forensic Psychiatry and Psychol 21(1):66–76

Porter S, ten Brinke L, Wallace B (2012) Secrets and lies: involuntary leakage in deceptive facial expressions as a function of emotional intensity. J Nonverbal Behav 36:23–37

Ruby CL, Brigham JC (1997) The usefulness of the criteria-based content analysis technique in distinguishing between truthful and fabricated allegations: a critical review. Psychol Publ Policy Law 3(4):705–737

Smith N (2001) Reading between the lines: An evaluation of the Scientific Content Analysis technique (SCAN). London, UK: Policing and Reducing Crime Unit: Police Research Series

ten Brinke L, Porter S, Baker A (2012) Darwin the detective: Observable facial muscle contractions reveal emotional high-stakes lies. *Evolution and Human Behavior*

The Global Deception Research Team (2006) A world of lies. J Cross-Cult Psychol 37(1):60–74

Undeutsch U (1989) The development of statement reality analysis. In: Yuille JC (ed) Credibility assessment. Kluwer Academic/Plenum Publishers, New York, pp 101–119

Vrij A (2007) Criteria based content analysis: a qualitative review of the first 37 studies. Psychol Publ Policy Law 11:3–41

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

Vrij A, Mann S (2006) Criteria-based content analysis: Am empirical test of its underlying process. Psychol Crime Law 12(4):337–349

Warren G, Schertler E, Bull P (2009) Detecting deception from emotional and unemotional cues. J Nonverbal Behav 33:59–69

Zaparniuk J, Yuille JC, Taylor S (1995) Assessing the credibility of true and false statements. Int J Law Psychiatry 18:343–352

Zuckerman M, DePaulo BM, Rosenthal R (1981) Verbal and nonverbal communication of deception. In: Berkowitz L (ed) Advances in experimental social psychology, vol 14. Academic Press, New York, pp 2–60