

# Testing the Effects of Nonverbal Behavior Training on Accuracy in Deception Detection with the Inclusion of a Bogus Training Control Group

Timothy R. Levine, Thomas Hugh Feeley, Steven A. McCornack, Mikayla Hughes, & Chad M. Harms

*Previous deception detection training studies have compared people receiving training in nonverbal behaviors associated with deception to control groups receiving no training and found that people who are trained are slightly to moderately more accurate than people who have not been trained. Recent research on the relationships between source veracity and specific nonverbal behaviors, however, suggests that those relationships are weak, inconsistent, and limited to high stakes lies. If specific nonverbal behaviors are not reliable indicators of deception, then one might wonder why training improves accuracy. This paper tests the speculation that the simple act of training, independent of the training content, may improve accuracy simply because those in training conditions process messages more critically. This speculation was tested in three experiments that included both no training and bogus training control groups. The bogus training group was most accurate in Study 1, but this finding failed to replicate in Study 2. A coding study (Study 3) examined behavioral differences in the stimulus tapes. The predicted differences were observed in a final experiment (Study 4) where training was based on the coded stimulus tapes. The results suggest that the effects of training are generally small and highly variable from message to message, that valid training does not produce large improvements over a bogus training control, and that bogus training can produce statistically significant improvements over a no-training control.*

---

Timothy Levine is Professor, Steven McCornack is Associate Professor, and Mikayla Hughes is a doctoral student in the Department of Communication at Michigan State University. Thomas Feeley is Associate Professor at Family Medicine Research Institute, and Department of Communication, State University of New York at Buffalo. Chad M. Harms is Assistant Professor in the Greenlee School of Journalism and Communication at Iowa State University. An earlier version of this paper was presented to the International Communication Association Annual Convention in Korea, May, 2002. Correspondence to Professor Timothy R. Levine, Department of Communication, Michigan State University, East Lansing, MI 48824-1212, USA; Tel.: 517 432 1124; Fax: 517 432 1192; email: levinet@msu.edu

*Keywords:* Lies; Deception; Detection; Training

Several previous studies have compared people receiving training in nonverbal behaviors associated with deception to control groups receiving no training (e.g., deTurck, Feeley, & Roman, 1997; deTurck, Harszrak, Bodhorn, & Texter, 1990; deTurck & Miller, 1990; Fieldler & Walka, 1993; Vrij, 1994; Vrij & Graham, 1997). This research implicitly or explicitly presumes the following: (a) a set of specific nonverbal behaviors exist that is diagnostically useful in distinguishing truthful messages from lies; (b) one of the reasons that research-naive people cannot accurately detect deception is because they do not rely on authentic deception cues and/or that they mistakenly rely on cues that have little diagnostic utility; and (c) peoples' judgmental accuracy would increase if they are trained to make veracity judgments based on authentic behaviors. Consistent with this reasoning, training studies have indeed found that people who are trained are slightly to moderately more accurate than people who have not been trained (e.g., deTurck et al., 1990; deTurck & Miller, 1990; Fieldler & Walka, 1993; Vrij & Graham, 1997), and a recent meta-analysis of these nonverbal training studies found that (although effects vary substantially from study to study) training, on average, increases detection accuracy by 4% (Frank & Feeley, 2003).

The most recent meta-analysis of the relationship between source veracity and specific nonverbal behaviors, however, suggests these relationships are weak, inconsistent, and limited to high motivation lies (DePaulo et al., 2003). If specific nonverbal behaviors are weak and unreliable indicators of deception, then one might question why nonverbal training improves accuracy. This paper argues that the simple act of training, independent of the training content, might improve accuracy simply because those in training conditions process messages more critically. This speculation is tested in three experiments that included both no training and bogus training control groups.

### **Deception Detection Accuracy**

Deception scholars agree that people's ability to distinguish truths from lies tends to be significantly, but only slightly, better than chance levels. Across studies, meta-analysis indicates that the mean accuracy rate is about 57% (Kraut, 1980), and literature reviews conclude that the accuracy rates reported in individual studies almost always fall within the range of 45–70% accuracy (e.g., Feeley & Young, 1998; Kalbfleisch, 1994; Miller & Stiff, 1993; Vrij, 2000). In short, the belief that deception detection accuracy rates are only slightly better than 50/50 is among the most well-documented and most commonly held conclusions in deception research.

Several reasonable explanations exist for peoples' relatively poor performance in deception detection. For example, systematic errors and biases in judgments such as the truth-bias are well documented (Levine, Park, & McCornack, 1999). Importantly for the current investigation, however, research-naive people focus on the wrong behaviors when trying to distinguish truths from lies (Miller & Stiff, 1993; Stiff & Miller, 1986). Although no 'sure-fire' deception cues exist, some statistically reliable correlates of deception have been reported in the literature (e.g., DePaulo et al., 2003;

Kraut, 1980; Zuckerman, DePaulo, & Rosenthal, 1981), as have several behaviors that research-naïve receivers tend to rely upon when making deception judgments (e.g., Stiff & Miller, 1986; Zuckerman, Koestner, & Driver, 1981). When comparing the 'authentic' deception cues to the behaviors that people tend to use, it becomes obvious that people are often influenced by some behaviors that lack predictive utility and people often ignore other diagnostically useful behaviors (Fieldler & Walka, 1993; Miller & Stiff, 1993; Stiff & Miller, 1986).

If there are reliable and diagnostically useful nonverbal behaviors associated with deception, and if one of the important reasons why people are inaccurate at detecting lies is because they focus on the wrong behaviors, then training people to look for authentic deception behaviors should lead to a substantial improvement in deception detection accuracy. Research on nonverbal training is consistent with this reasoning. Those receiving nonverbal training in nonverbal behaviors believed to have diagnostic utility have been found to be significantly more accurate than people in no training, control groups (e.g., deTurck et al., 1990; deTurck & Miller, 1990; Fieldler & Walka, 1993; Frank & Feeley, 2003; Vrij, 1994). Nevertheless, the across-study average effect size is modest ( $r = .20$ ), findings vary significantly and substantially from study to study, and important methodological limitations are commonplace (Frank & Feeley, 2003).

### **Nonverbal Deception Cues**

It is commonly believed that truth-tellers and liars exhibit different patterns of nonverbal behaviors. For example, in an often cited meta-analysis, Zuckerman, DePaulo, & Rosenthal, (1981) found that relative to truth-tellers, liars used more adaptors, made more speech errors, paused more often, and had higher pitch. As another example, deTurck and Miller (1985) reported that liars used more adaptors, more hand gestures, more speech errors, more pauses, longer response latencies, and shorter talk duration than truth-tellers. Drawing upon these findings and others, literature reviews often conclude that specific nonverbal behaviors are diagnostically useful indicators of deception (e.g., Burgoon, Buller, & Woodall, 1989; Miller & Stiff, 1993).

These conclusions, however, have been recently challenged (e.g., DePaulo et al., 2003; Frank & Feeley, 2003; McCornack, 1997). Some scholars argue that little reason exists to expect behavioral differences between everyday, low consequence truths and lies (DePaulo, Kashy, Kirkendol, & Epstein, 1996; McCornack, 1997) and that marked behavioral differences may be limited to high stakes lies (Frank & Feeley, 2003). Consistent with this reasoning, the most recent and most thorough meta-analysis of deception-related behaviors indicates that the effects of source veracity on specific source behaviors are often small and unreliable (DePaulo et al., 2003). For example, of the deception cues listed above, only talk duration and pitch were significantly related to source honesty across studies, and the effects for these two behavior types were small ( $r < .20$ ; DePaulo et al., 2003).<sup>1</sup> Motivation to deceive, however, was found to be an important moderator. Larger behavioral differences between truth-tellers and liars are observed as the stakes increase.

## **Current Research Focus**

Recent findings concerning the lack of substantial nonverbal differences present an interesting paradox for nonverbal training studies. If the nonverbal behaviors that previous researchers have used to train judges really have little diagnostic utility, then why does training appear to improve accuracy? We propose that training effects may have very little to do with which cues judges are actually trained to consider.

One of the most well-documented reasons people are inaccurate is because they engage in mindless decision rules that might lead to a truth-bias (Levine et al., 1999). Judges who are trained in specific nonverbal behavioral cues may be more accurate simply because they are paying more attention to source behaviors. Increased attention to source behaviors coupled with increased awareness of the possibility of deception should reduce mindless processing and truth-bias thereby increasing accuracy. To test this reasoning, however, we need the nonverbal training equivalent of a placebo control group.

In the three experiments reported below, participants are randomly assigned to one of three experimental groups. One group of judges receives brief training in four nonverbal behaviors that have (or are thought to have) diagnostic utility. A second group is given no training at all. A third group is trained in a manner identical to the first group, except that the specific behaviors taught are unrelated to deception, and hence have no diagnostic utility. To the extent the traditional thinking about nonverbal training is correct, we would expect that the valid training group would be more accurate than the no-training control group, and the bogus training control would be least accurate (or similar to the training control). If the alternative reasoning is correct, however, it is expected that both the valid training and the bogus training groups yield significantly higher accuracy rates than the no-training control.

## **Study 1**

### *Method*

#### *Stimulus materials*

Two undergraduate students (one male and one female) from an upper division Speech class at the University of Hawaii served as message sources in exchange for extra credit. A videotape containing a series of truthful and deceptive statements was made for each source. First, the source completed an attitude survey containing 12 Machiavellianism scale items (Christie & Geis, 1970). Each item used a 10-point response format. After completing each of the 12 items, the experimenter changed half of the items (selected at random) to an answer that was five points different (half the scale length) from the subject's true answer. The source was then told to report his or her true answer for half of the 12 items. For the remaining (changed) items, the source was told to report the answer which was five points (half the scale length) different from his or her true answer. The source was then seated in a chair and told to state his or her response to each item, and to briefly explain each answer. This was videotaped. The camera was placed so that the source's entire body was visible. This

procedure resulted in the creation of two (one male source and one female source) 'master tapes,' each containing 12 statements, 6 true and 6 false.

Each master tape was digitized and edited with an AVID computer video editing system. First, 3-second pauses were placed between each of the 12 segments. Second, four segments were deleted from each tape. Two randomly selected honest statements and two randomly selected lies were removed to form each source to create a stimulus tape with 16 messages. The tape contained 8 messages (4 true, 4 false) from the male source and 8 messages (4 true, 4 false) from the female source. Other combinations of deleted items were used to make other tapes, but these other stimulus tapes were not used in the present investigation.

Two training tapes were also made. On each tape, a male professor who is a well-known deception researcher, but was not involved with the current data collection, gave a five-minute lecture on the nonverbal behaviors associated with deception. On one tape, it was explained that liars relative to truth-tellers exhibit longer response latencies, more adaptors, more speech errors, and more pauses. As previous research has linked each of these behaviors with deception (e.g., deTurck & Miller, 1985; Zuckerman, DePaulo, & Rosenthal, 1981), this tape was used in the *valid training* condition. On the second tape, it was explained that recent research has shown that liars exhibit less eye contact, talk faster, engage in more posture shifts, and have more foot movements than truth-tellers. These specific behaviors were chosen because previous research has found that they are not linked to deception, yet these might seem to be plausible deception cues for research-naïve participants. This second tape constituted the *bogus training* tape. On both tapes, the relevant nonverbal behaviors were defined and explained, and examples were given. Viewers were instructed to rely on the behaviors described when making truth-lie judgments. Effort was made to keep the two tapes identical, except for content related to the four nonverbal behaviors.

### *Participants*

The participants were 256 (82 males and 174 females) undergraduate students enrolled in communication classes at the Michigan State University. The participants ranged in age from 18 to 25 ( $M = 19.91$ ,  $SD = 1.36$ ). All received extra credit in exchange for their participation.

### *Design and procedures*

The design was a one-way, three independent groups quasi-experiment with a valid training group ( $n = 71$ ), a bogus training control ( $n = 61$ ), and a no-training control ( $n = 124$ ). In the valid training condition, participants were first shown the valid training tape. Participants in the bogus training group watched the bogus training video. In the no-training control, participants saw neither training tape, and went straight from instructions to the message judgment task.

Experimental sessions were run in groups during regular class time. Each group was randomly assigned to one of the three experimental conditions. The respondents were

told that they would be watching a series of videotaped segments in which another student was discussing his or her answers to an attitude survey. If the participants were in one of the two training conditions, they saw a training tape at this point. The participants were then told that they were to judge whether each message was a truth or lie. After the instructions were completed, the participants were shown each video clip one at a time. After viewing each of the 16 videotaped segments, the tape was paused, and the participants were asked to make a dichotomous 'truth/lie' judgment that served as a measure of accuracy. Judgmental accuracy was computed as the proportion of accurate judgments to total judgments across the 16 segments (0–100% accuracy). The participants also completed demographic items and were debriefed.

### *Results and Discussion*

The data were analyzed with a one-way independent groups ANOVA with accuracy as the dependent variable. Tests of statistical assumptions revealed that scores on the dependent measure were approximately normally distributed and that the assumption of homogeneity of variance was met. The omnibus  $F$ -test was statistically significant,  $F(2, 252) = 7.47, p < .001, \eta^2 = .06$ . Contrary to expectations, participants were most accurate in the bogus training condition ( $M = 55.7\%, SD = 10.1\%$ ), followed by the no-training control group ( $M = 52.1\%, SD = 11.4\%$ ), and the valid training conditions ( $M = 48.6\%, SD = 9.3\%$ ). Follow-up  $t$ -tests showed that each pair of the means was significantly different, but only the bogus training and valid training conditions differed significantly with more conservative Scheffe tests.

These results were not anticipated, and why training in supposedly authentic nonverbal behavior would lead to significantly poorer detection rates than training bogus cues is unclear. Perhaps, inconsistent with previous research, the sources on the stimulus tape may have actually exhibited shorter response latencies, fewer adaptors, fewer speech errors, fewer pauses, less eye contact, faster talking, more posture shifts, and more foot movements when lying than when telling the truth. That is, the bogus behaviors may have had diagnostic utility, and the presumably valid behaviors may have lacked utility in stimulus tapes. Due to the unexpected findings, a replication was conducted.

## **Study 2**

### *Method*

The participants were 90 (26 males and 64 females) undergraduate students enrolled in communication classes at Michigan State University. The participants ranged in age from 19 to 29 ( $M = 21.51, SD = 1.49$ ). All received extra credit in exchange for their participation. The design was again a one-way, three independent groups experiment with participants randomly assigned to a valid training group ( $n = 28$ ), a bogus training control ( $n = 31$ ), or a no-training control ( $n = 31$ ). The procedures and measurement were identical to Study 1 except that experimental sessions were run in groups of less

than 10 individuals outside of regular class time. Thus, unlike Study 1 that used intact groups, Study 2 was a true experiment with random assignment.

### *Results and Discussion*

The data were analyzed with a one-way independent groups ANOVA, with accuracy as the dependent variable. Tests of statistical assumptions revealed that scores on the dependent measurement closely approximated normality and that the assumption of homogeneity of variance was met. The omnibus *F*-test was not statistically significant,  $F(2, 87) = 0.09, p = ns, \eta^2 = .00$ . Participants were most accurate in the no-training condition ( $M = 53.2\%, SD = 12.7\%$ ), closely followed by the valid training group ( $M = 52.9\%, SD = 09.2\%$ ), and the bogus training condition ( $M = 52.0\%, SD = 11.1\%$ ). Follow-up *t*-tests showed that none of the means were significantly different. The statistical power to detect the size of effect found in Study 1 was approximately .50.

Whereas the data do not replicate the results of Study 1, the data are not inconsistent with the findings of Study 1 either. Calculating confidence intervals around the cell means in Studies 1 and 2 shows that in each case the confidence intervals overlap. Thus, sampling error and differences in statistical power cannot be ruled out as an explanation for the differences in results.

To further explore what might be happening in the data, the results of both experiments were reanalyzed with message as a repeated factor in  $3 \times 16$  mixed ANOVAs. Substantial message effects [Study 1,  $F(15, 3780) = 94.37, p < .05, \eta^2 = .25$ ; Study 2,  $F(15, 1305) = 28.77, p < .05, \eta^2 = .23$ ] and weaker but statistically significant message by training interactions [Study 1,  $F(30, 3780) = 7.70, p < .05, \eta^2 = .04$ ; Study 2,  $F(30, 1305) = 3.08, p < .05, \eta^2 = .05$ ] were evident.

Accuracy rates for individual messages in the different training conditions are presented in Table 1. Examination of the training effects within messages suggests that the results of the two experiments are consistent for some messages but not others, and that training had inconsistent effects from message to message. For example, participants were substantially more accurate in the no-training control than in either training condition for the first message. Conversely, on messages 8 and 11, participants were substantially more accurate in the two training conditions than in the no-training control. No consistent patterns across both messages and experiments are evident. To further explore these data, specific nonverbal behaviors were assessed in Study 3.

## **Study 3**

### *Method*

#### *Participants*

The participants were 96 (25 males and 71 females) undergraduate students enrolled in communication classes at Michigan State University. The participants ranged in age from 18 to 26 ( $M = 21.07, SD = 1.25$ ). All received extra credit in exchange for their participation.

**Table 1** Accuracy by Study, Training Condition, and Message

Message	Study 1			Study 2			R
	C	B	V	C	B	V	
Male source							
1 (Truth) <sub>1,2</sub>	.71	.32	.17	.61	.22	.17	R
2 (Lie)	.19	.18	.15	.48	.19	.36	
3 (Truth) <sub>1</sub>	.48	.78	.42	.81	.68	.61	
4 (Truth) <sub>1,2</sub>	.80	.78	.42	.55	.81	1.00	
5 (Lie) <sub>1</sub>	.22	.42	.27	.29	.35	.29	r
6 (Lie) <sub>1</sub>	.52	.42	.27	.35	.45	.43	
7 (Truth) <sub>1</sub>	.23	.12	.10	.16	.13	.04	r
8 (Lie) <sub>1,2</sub>	.52	.72	.73	.25	.61	.57	R
Female source							
9 (Lie) <sub>1</sub>	.17	.03	.04	.13	.09	.00	r
10 (Lie) <sub>1</sub>	.58	.83	.70	.58	.81	.71	r
11 (Truth) <sub>1,2</sub>	.54	.85	.87	.45	.74	.89	R
12 (Truth)	.82	.85	.87	.87	.77	.86	r
13 (Truth)	.85	.92	.80	.81	.77	.82	r
14 (Lie) <sub>1</sub>	.35	.45	.20	.52	.45	.42	
15 (Truth) <sub>2</sub>	.85	.85	.87	.94	.65	.79	
16 (Lie)	.51	.40	.42	.71	.58	.50	
Total <sub>1</sub>	.52	.56	.49	.53	.52	.52	

*Note.* 1 denotes a significant effect for training condition in Study 1, 2 marks a significant training effect in Study 2, an R denotes a significant replication, and an r marks a similar pattern in means. C = control group, B = bogus training group, and V = valid training group.

### *Design and procedures*

The study was an eight independent groups design with different groups rating one of eight different nonverbal behaviors. The procedures were similar to the no-training control conditions in Study 2 except that participants were instructed to count or rate a specific behavior instead of making veracity judgments. The eight behaviors assessed corresponded to the behaviors included in the valid and bogus training tapes. Eye contact, speech rate, pausing, and response latencies were rated on a 5-point scale, while instances of posture shifts, foot movements, adaptors, and speech errors were counted.

### *Results and Discussion*

The data were initially analyzed with eight paired *t*-tests, one for each behavior. Four of the eight tests produced statistically significant results, although the effect sizes were uniformly small. Compared to truthful messages, the lies exhibited shorter response latencies, fewer speech errors, fewer pauses, and fewer foot movements. Thus, three out



**Table 2** Differences in Nonverbal Behaviors Between Truths and Lies

	Truth	Lie	<i>t</i>	<i>p</i> <
4 "valid" behaviors				
Response latencies	2.41	2.03	-3.95	.002
Speech errors	3.81	3.35	-3.07	.011
Adaptors	2.33	2.33	0.00	—
Pauses	2.35	2.05	-3.35	.007
4 "bogus" behaviors				
Fast talking	2.91	2.94	0.69	.504
Posture shifts	1.22	1.32	0.74	.478
Eye gaze	2.86	2.74	-1.46	.172
Foot movements	1.34	1.09	-2.87	.015

Note. Negative *t*-values indicate findings opposite to what was taught in the training videos.

of the four valid nonverbal behaviors were in the direction opposite to what would be expected from previous research. The bogus behaviors, however, proved to be actually bogus. Three did not significantly differentiate between truths and lies, while foot movements functioned in the opposite direction to what was claimed in the bogus training tape. The complete results are presented in Table 2.

Recall that the exploratory analyses reported above found that training had highly variable effects from message to message. These findings were further explored with a series of 8 (one for each coded behavior)  $\times$  2 (message source)  $\times$  2 (veracity: truth vs. lie)  $\times$  4 (message instantiation) fully repeated ANOVAs. The results of these analyses are presented in Table 3.

Consistent with DePaulo et al. (2003), the results indicate that the effects for source honesty on specific nonverbal behaviors are small (i.e.,  $\eta^2 \leq .03$ ). Message source and

**Table 3** Effect Sizes for Statistically Significant Main Effects and Interactions for Message Source, Source Veracity, and Message Instantiation on Nonverbal Behaviors

	Source of variation in ANOVA						
	S	V	M	S $\times$ V	S $\times$ M	V $\times$ M	S $\times$ V $\times$ M
Response latencies	.18	.03	.14	.15	.04	.06	.04
Speech errors	.65	.01	—	.05	.03	.04	.03
Adaptors	.09	—	.15	.06	.17	.05	.05
Pauses	.23	.02	.13	.10	.12	.04	.05
Fast talking	—	—	—	.05	.11	—	.08
Posture shifts	.13	—	—	—	—	.05	.04
Eye gaze	.40	—	.04	.05	.10	.03	.05
Foot movements	.04	.01	—	.01	.04	.21	.08

Note. Values in the table are eta squares. S refers to message source, V to veracity, and M to message.

message instantiation within source contributed substantial variability to observed behaviors, and the message veracity effects are moderated by two- and three-way interactions with source and message instantiations. These source and message effects likely explain why the effects of training were variable across messages.

The results shed light on the unanticipated findings obtained in Study 1. Recall that Study 1 found that participants in the valid training condition were less accurate than participants in the bogus training condition and in the no-training control group. Given the results of Study 3, a plausible explanation for this finding is that the behaviors trained in the valid condition were actually associated with honesty in the stimulus tapes used. These findings suggest that if people are trained in ‘wrong direction’ behaviors, training can be counterproductive. If the valid behaviors in Study 1 actually had diagnostic utility, the data might have come out as originally expected. To test this reasoning, a fourth study was conducted where the content of the nonverbal training was based on the coding of the stimulus tapes from Study 3.

## **Study 4**

### *Method*

The participants were 158 (86 males and 72 females) undergraduate students enrolled in communication classes at the same University. The participants ranged in age from 17 to 23 ( $M = 19.83$ ,  $SD = 1.26$ ). All received extra credit in exchange for their participation.

The design was again a one-way, three independent groups experiment with randomly assignment to a valid training group ( $n = 52$ ), a bogus training control ( $n = 52$ ), or a no-training control ( $n = 54$ ). The procedures and measurement were identical to Study 2 except that the content of the training inductions was changed to reflect the results of Study 3. In the valid training condition, participants were told that liars would exhibit shorter response latencies, fewer speech errors, less pausing, and less foot movements than truth-tellers. Each of these behaviors was shown to have diagnostic utility in Study 3. In the bogus training condition, participants were instructed to look for a lack of eye contact, more adaptors, faster talking, and increased posture shifts. These behaviors were found to have little utility in Study 3. Otherwise, the same scripts were used and the content for the training tapes were as similar to the previous tapes as possible.

### *Results and Discussion*

The data were analyzed with a series of one-way independent groups ANOVAs with total accuracy, truth and lie accuracy, and truth-bias as the dependent variables. Tests of statistical assumptions revealed that the assumption of homogeneity of variance was met. Scores on truth accuracy and lie accuracy closely approximated normality. The distribution of total accuracy was more leptokurtic than normal and truth-bias scores were positively skewed.

The omnibus  $F$ -test for detection accuracy was statistically significant,  $F(2, 157) = 7.92, p < .01, \eta^2 = .09$ . Participants were most accurate in the valid training condition ( $M = 58.2\%$ ,  $SD = 11.8\%$ ), followed by the bogus training group ( $M = 55.9\%$ ,  $SD = 10.2\%$ ), and the no-training condition ( $M = 50.2\%$ ,  $SD = 9.7\%$ ). Follow-up Scheffe tests showed that both the valid training and the bogus training were significantly greater than the no-training control, but did not differ from each other.

Statistically significant differences were also observed for truth-bias,  $F(2, 157) = 12.39, p < .01, \eta^2 = .14$ . Participants were most truth-biased in the no-training control ( $M = 68.5\%$ ,  $SD = 14.0\%$ ), followed by the valid training group ( $M = 59.6\%$ ,  $SD = 10.8\%$ ), and the bogus training condition ( $M = 56.6\%$ ,  $SD = 13.4\%$ ). Follow-up Scheffe tests showed that both the valid training and the bogus training were significantly lower than the no-training control, and did not differ from each other.

Following Levine et al. (1999), accuracy rates were calculated separately for truths and lies. Statistically significant differences were observed for lie accuracy,  $F(2, 157) = 18.96, p < .01, \eta^2 = .20$ , but not truth accuracy,  $F(2, 157) = 2.15, p = .12, \eta^2 = .03$ . Participants were significantly better at correctly identifying lies in the bogus training ( $M = 49.3\%$ ) and valid training ( $M = 48.6\%$ ) conditions than in the no-training control ( $M = 31.7\%$ ). Cell means are presented in Table 4. The statistical power to detect the medium effects ( $d = .50$ ) between individual cell means was approximately .71.

The results of the three training experiments (i.e., Studies 1, 2, and 4) are summarized in Table 5. Within condition detection accuracy in the bogus training and the no-training control conditions did not vary across the studies and there was substantial overlap in the 95% confidence intervals with conditions but across studies. The results for the valid training, however, differed significantly in Studies 1 and 3. The 95% confidence interval around the valid training mean in Study 1 was 46.4–50.8%. In Study 3, the 95% CI was 54.5–61.4%. The 95% CI around the mean for the valid training condition in Study 2, 49.5–56.3%, overlaps with the CIs in both Studies 1 and 4.

## General Discussion

The series of studies reported here examined the effects of nonverbal training on deception detection accuracy. The key difference between these studies and previous research

**Table 4** Detection Accuracy by Condition in Study 4

	Condition		
	Valid training	Bogus training	No training
Total accuracy	58.17% <sub>a</sub>	55.89% <sub>a</sub>	50.23% <sub>b</sub>
Truth accuracy	67.79%	62.50%	68.75%
Lie accuracy	48.56% <sub>a</sub>	49.28% <sub>a</sub>	31.71% <sub>b</sub>
Truth-bias	59.62% <sub>a</sub>	56.61% <sub>a</sub>	68.52% <sub>b</sub>

*Note.* Means with different subscripts differ significantly at  $p < .05$ .

**Table 5** Detection Accuracy by Training Condition

	Condition		
	Valid training	Bogus training	No training
Study 1	48.59% <sub>a</sub>	55.73%	52.07%
Study 2	52.90%	52.02%	53.23%
Study 4	58.17% <sub>a</sub>	55.89%	50.23%

*Note.* Means with the same subscripts differ significantly at  $p < .05$ .

on training was the addition of a bogus training control group. Previous training studies show that, when compared to those receiving no training at all, training can improve accuracy. These previous studies, however, confounded specific training content with demand effects and heightened vigilance unrelated to specific nonverbal behaviors. This led to the novel prediction that even training people in cues with no lack diagnostic utility might lead to increased accuracy.

The data were mostly consistent with this reasoning. Even though coding in Study 3 confirmed that the behaviors in the bogus condition lacked diagnostic utility, participants in this condition were more accurate than those in the no-training control in Studies 1 and 4. In Study 4 where the valid training cues had diagnostic utility, both the valid and the bogus training conditions resulted in improved accuracy over the no-training control. However, consistent with the traditional line of reasoning, participants in the valid training condition in Study 1 were less accurate than those in the other conditions presumably because the cues they were trained to look for were actually associated with honesty.

Together these results suggest that both the traditional training explanation and the demand effects hypothesis may have merit. The traditional training explanation, but not the demand effects hypothesis, explains the decrease in accuracy in the valid training condition in Study 1 where participants were inadvertently trained to look for cues that were actually associated with honesty. The demand effects explanation, but not the traditional view, however, accounts for the improvement in the bogus training over no-training controls in Studies 1 and 4.

These conclusions, however, are somewhat tenuous since no significant differences were found in Study 2. The failure to replicate may be attributable to lower statistical power, but that of course cannot be known. Whereas the patterns observed in Studies 1 and 4 were not replicated in Study 2, neither were the results of Study 2 significantly discrepant from Studies 1 and 4.

There are at least three important implications of these results. First, the inclusion of a bogus training control group represents an important design improvement over previous studies, and future researchers studying nonverbal training are encouraged to include such a condition. The results of the current research that training participants to look for behaviors with no diagnostic value is sufficient to produce gains in accuracy commensurate with the gains observed in the previous research (cf. Frank & Feeley,

2003). Thus, strong evidence that some specific training procedure leads to improved accuracy rates requires the conceptual equivalent of a placebo training control.

Second, the most striking feature of the current results was the high degree of variability between the two sources and from message to message within source and veracity condition. For training to have efficacy in detecting whether a specific message from a specific person is honest or not, the behaviors trained would need to have diagnostic utility that is general across messages, people, and situations. The present data suggest that specific behaviors were highly variable across the two sources and the eight specific messages. Thus, relying on specific nonverbal indicators may improve accuracy rates over a large number of trials across a range of situations, but specific behaviors may have little value when examining single messages from individual sources.

Third, and as a consequence, a major challenge facing those who might use nonverbal training in applied contexts is the determination of which behaviors to include in the training. As mentioned in the literature review, previous research is not in agreement about which behaviors have diagnostic utility (cf. DePaulo et al., 2003; deTurck & Miller, 1985; Fiedler & Walka, 1993; Zuckerman, dePaulo, & Rosenthal, 1981). For example, even though participants in the valid training condition in Studies 1 and 2 were trained in cues that previous research suggested were most likely valid, those cues proved to be indicators of honesty rather than deceit in the particular messages in the current experiments. Practitioners, however, cannot know with certainty whether or not a particular statement is deceptive or not, nor can they know which specific behaviors, if any, have diagnostic utility in the messages under consideration.

### *Limitations and Future Directions*

Several limitations to this study could be addressed in future research on lie detection training. First, the stimulus materials did not include different types of lies and different motivations for lying. We used lies about private feelings/attitudes and there were no negative consequences for getting caught in a lie. Following Frank and Feeley (2003), future research could employ high stakes lies about actions and facts. These are the types of lies that are most important to detect, and they are also the types of lies where nonverbal differences should be maximized. Although high stakes lies were not needed to test our bogus training hypothesis, the current design did not provide a strong test of genuine training effects.

Second, the current design was limited to two senders and 16 messages. Adding additional senders would enhance surface generalizability. The true litmus test of training is its ability to raise lie detection across several different senders telling several different types of lies (and truths). A more diverse set of stimulus materials would provide a riskier and more satisfying test of training effectiveness.

Third, the training inductions were also rather limited. Frank and Feeley (2003), for example, recommend that training should be more than one hour and should include both nonverbal cue information and a feedback session. As with the high stakes lies issue, the current procedures were deemed adequate to test our bogus training hypothesis, but more extensive training would be needed to provide a strong test of genuine

training effects. Further, it should be noted that the training in most previous training studies is similar to the current studies. Therefore, extensiveness of training is not a viable explanation of the difference between the current results and previous findings. Nevertheless, Frank and Feeley's (2003) recommendations for more extensive training inductions have merit.

Finally, accuracy was assessed immediately after the training intervention. To our knowledge, there are no data available on the long term effectiveness of nonverbal training. Future research might examine the effects of training over time.

### *Summary*

The studies presented here tested the effects of nonverbal training with the inclusion of a placebo control group. This constitutes a substantial methodological improvement over previous studies, and it is hoped that future research adopts this innovation.

It was predicted that training participants in nonverbal behaviors that lacked diagnostic utility would nevertheless improve detection accuracy. The data in Studies 1 and 4 were consistent with this, but the results of Study 2 were equivocal. Although tentative, the results are encouraging and future researchers might consider the possibility of placebo effects when evaluating the effectiveness of deception detection training programs.

The current findings also provided an interesting twist of previous training results. Previous studies found that training can produce modest improvements in detection accuracy presumably because attention is paid to cues with diagnostic utility. In Study 1, the valid training condition actually had participants watch for counterproductive cues (i.e., the behaviors that were said to be indicative of deceit were actually associated with honesty). Consistent with the logic of nonverbal training, detection accuracy declined. These data suggest that training programs may have intended effects, unintended effects, and placebo effects, and that these may be additive. Again, caution is advised because the results did not replicate in Study 2.

Perhaps the most important findings, however, concern the observation of substantial variability in nonverbal behaviors within truths and lies. Variability attributable to the message source, the message instantiation, and statistical interactions among the factors swamped the effects of message veracity in the current data. This variability creates a serious challenge for the effectiveness of nonverbal deception detection training. For training programs to be theoretically viable and practically efficacious, trainers must know a priori what behaviors have diagnostic utility. To the extent that nonverbal behaviors are inconsistently displayed across messages, people, and contexts, the specific behaviors trained will have limited diagnostic utility, and training effectiveness will be limited accordingly.

### **Note**

- [1] Whereas some behaviors might show statistically significant differences between truth and lies, to the extent that the effect size for the difference is small, the utility of that behavior for

detection accuracy training might be questioned. Trends that are statistically reliable at  $p < .05$  are not necessarily useful in this context, and a correlation of  $r = .10$  might not be visible even to the trained eye.

## References

- Burgoon, J. D., Buller, D. B., & Woodall, W. G. (1989). *Nonverbal communication: The unspoken dialogue*. New York: Harper & Row.
- Christie, R., & Geis, F. L. (1970). *Studies in Machiavellianism*. New York: Academic Press.
- DePaulo, B. M., Kashy, D. A., Kirkendol, M. M., & Epstein, J. A. (1996). Lying in everyday life. *Journal of Personality and Social Psychology, 70*, 979–995.
- DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin, 129*, 74–118.
- deTurck, M. A., Feeley, T. H., & Roman, L. (1997). Visual and vocal cue training in behavioral lie detection. *Communication Research Reports, 14*, 249–259.
- deTurck, M. A., Harszrak, J. J., Bodhorn, D. J., & Texter, L. A. (1990). The effects of training social perceivers to detect deception from behavioral cues. *Communication Quarterly, 38*, 189–199.
- deTurck, M. A., & Miller, G. R. (1985). Deception and arousal: Isolating the behavioral correlates of deception. *Human Communication Research, 12*, 181–201.
- deTurck, M. A., & Miller, G. R. (1990). Training observers to detect deception: Effects of self-monitoring and rehearsal. *Human Communication Research, 16*, 603–620.
- Feeley, T. H., & Young, M. J. (1998). Humans as lie detectors: Some more second thoughts. *Communication Quarterly, 46*, 109–126.
- Fiedler, K., & Walka, I. (1993). Training lie detectors to use nonverbal cues instead of global heuristics. *Human Communication Research, 20*, 199–223.
- Frank, M. G., & Feeley, T. H. (2003). To catch a liar: Challenges for research in lie detection training. *Journal of Applied Communication Research, 31*, 58–75.
- Kalbfleisch, P. J. (1994). The language of detecting deceit. *Journal of Language and Social Psychology, 13*, 469–496.
- Kraut, R. (1980). Humans as lie detectors. *Journal of Communication, 30*, 209–216.
- Levine, T. R., Park, H. S., & McCornack, S. A. (1999). Accuracy in detecting truths and lies: Documenting the 'veracity effect.' *Communication Monographs, 66*, 125–144.
- McCornack, S. A. (1997). The generation of deceptive messages: Laying the groundwork for a viable theory of interpersonal deception. In J. O. Greene (Ed.), *Messages production*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Miller, G. R., & Stiff, J. B. (1993). *Deceptive communication*. Newbury Park, CA: Sage.
- Stiff, J. B., & Miller, G. R. (1986). 'Come to think of it . . .': Interrogative probes, deceptive communication, and deception detection. *Human Communication Research, 12*, 339–357.
- Vrij, A. (1994). The impact of information and setting on detection of deception by police detectives. *Journal of Nonverbal Behavior, 18*, 117–136.
- Vrij, A. (2000). *Detecting lies and deceit: The psychology of lying and the implications for professional practice*. Chichester, UK: John Wiley & Sons.
- Vrij, A., & Graham, S. (1997). Individual differences between liars and the ability to detect lies. *Expert Evidence: The International Digest of Human Behavior Science and Law, 5*, 144–148.
- 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. 1–59). New York: Academic Press.
- Zuckerman, M., Koestner, & Driver, R. (1981). Beliefs about cues associated with deception. *Journal of Nonverbal Behavior, 6*, 105–115.