To Catch a Liar: Challenges for Research in Lie Detection Training

Mark G. Frank and Thomas Hugh Feeley

ABSTRACT   *Can we train people to detect deception? It is the contention of this article that communication scholars should learn how to train law enforcement professionals on how to detect high stake lies, like those faced by police, judges, customs officials, immigration officials, and so forth. It is proposed that in order to know whether we can train or should bother to train people to detect deception, each training study must meet 6 challenges: (1) relevance, (2) high stakes, (3) proper training, (4) proper testing, (5) generalizability across situations, and (6) generalizability over time. Our quantitative review of the literature suggests that training does significantly raise lie detection accuracy rates. Meta-analytic findings indicate a mean effect size of r — .20 across 20 (11 published studies) paired comparisons of lie detection training versus the control group (i.e., those without some type of training). It should be noted that the majority of the studies that attempt to train lie detectors fall short on many of the above challenges. Current research in lie detection training may actually underestimate the ability to train lie detectors due to the stimulus materials employed in most experiments.*
KEY WORDS: lie detection, lying, deception, training, veracity.

O ne of the most critical, life or death judgments law enforcement professionals can make is whether someone is lying or telling the truth. A murderer is free to kill again if an investigating police officer believes his or her lies. The same applies to a jury or judge who believes a lying witness. Terrorists can smuggle in weapons of death and destruction if a customs or immigration official believes his or her lies. A particularly potent recent example is the case of Cary Stayner, who was able to lie convincingly to police investigators in their initial search for the murderer of three women in Yosemite National Park. Upon release, Stayner killed another woman before he was caught. Conversely, disbelieving a truthful person can have equally horrifying consequences. Innocent men and women have been tried and convicted of crimes, and forced to spend years in prison or even be executed, whereas the real perpetrator goes free (Dwyer,

Mark G. Frank (Ph.D., Cornell University) is Associate Professor in the School of Communication, Information, and Library Sciences at Rutgers State University of New Jersey and Thomas Feeley (Ph.D., State University of New York at Buffalo) is Research Assistant Professor and Director of Educational Development in the Department of Family Medicine at State University of New York at Buffalo. This article was supported by grants from the Rutgers University Research Council and the Federal Bureau of Alcohol, Tobacco, & Firearms to the first author. A previous version of this paper was presented to the National Communication Association annual convention, Seattle, WA, 2000. All correspondence should be directed to the first author at mgfrank@scils.rutgers.edu.

Neufeld, & Scheck, 2000). A recent case in New Jersey found a man released from prison when a court-ordered DNA test ruled him out as the rapist of a college student; unfortunately, this man had spent 9 years in prison based partly on the investigators' belief that he was lying about his whereabouts during the rape. These are but a fraction of the life or death high stake judgments faced by professional lie catchers.

Although an integral part of their job, the evidence from laboratory studies suggests that most professional lie catchers are not much better than laypeople at distinguishing truths from lies (e.g., DePaulo & Pfiefer, 1986; Ekman & O'Sullivan, 1991; Ekman, O'Sullivan, & Frank, 1999; Garrido & Mausip, 1999; Kraut & Poe, 1980). However, there is some evidence that some professional lie catchers do outperform laypeople. Ekman and O'Sullivan (1991) found that the U.S. Secret Service scored significantly above chance in distinguishing nursing students who saw relaxing movies from those who saw disturbing movies. Ekman, O'Sullivan, and Frank (1999) found that particularly motivated police and judges were able to outperform their less motivated counterparts when distinguishing truths and lies about opinions. Observers in the Ekman et al. study judged the veracity of students who had much to gain from a successful lie (and a successful truth) and much to lose if judged as deceptive, regardless of actual veracity (see Frank & Ekman, 1997). It appears that the reason these two studies found evidence for high rates of lie detection accuracy for professionals, when others had not, was that these studies featured liars who were experiencing strong emotions due to the stakes associated with the success or failure of their lies. Moreover, both these studies showed that there were clues that distinguished the liars and the truth tellers at rates of at least 80% accuracy. Thus, the high rates of accuracy for these professionals suggest that they are picking up on clues in the behavior of the liars/truthtellers that are guiding their judgments. Perhaps if there are such diagnostic clues to lying, then these clues can be taught to others, thus improving the accuracy of lie detection.

Given the importance of distinguishing between truths and lies, it is surprising that police officers tend to receive little training in deception detection (Bull, 1989). The first author has trained over 1,000 law enforcement personnel, and estimates that fewer than 10% of these people have received any specific training on detecting lies. Moreover, many of the 10% that have received training were given an oversimplified set of cues that have not been substantiated by empirical research (e.g., if a suspect touches her nose or crosses her arms she is lying). This information is often delivered in pamphlets or magazine articles, and not substantiated with empirical research articles.

We suggest a number of reasons why law enforcement personnel have received little training in lie detection. First, the deception literature suggests that there are no perfect clues to lying that would allow for foolproof training (e.g., DePaulo, Stone, & Lassiter, 1985; Ekman, 1992; Knapp & Comadena, 1979; Zuckerman, DePaulo, & Rosenthal, 1981). Thus it is understandable why police would be reluctant to train without these definitive clues. Second, in conjunction with the first point, those clues that research has shown to be associated with lying are only probabilistically related (DePaulo, 1994). A clue that indicates lying 70% of the time within a particular context means that an investigator who relies upon this clue within that context will be wrong 30% of the time. He or she will also miss those lies that occur when that clue is not present. This too does not seem specific

enough for many professionals. Third, there is an aversion in law enforcement to the academics presenting these probabilistic findings. Fourth, many police officers argue that the base rate for lying in encounters with suspected individuals is as high as 75%. If we assume for a moment that this is true—without any scientifically defensible evidence that it is so—then an officer who adopts a decision rule that all people are lying will be correct 75% of the time. An officer who is accurate 75% of the time (or who believes that they are correct at a higher rate than that) would not feel the need to seek training on lie detection. Evidence shows that police become more confident that they are correct with on the job experience, without any corresponding increase in accuracy (DePaulo & Pfiefer, 1986).

These arguments suggest why law enforcement and other professional lie catchers would not want to subject themselves to extensive lie detection training. What we are interested in is whether these arguments are justified in light of the scientific evidence for or against our ability to improve a professional's ability to distinguish lies from truths and whether communication researchers can offer insight into this process.

Because of the weak relationship between various behavioral clues and lying described above, it is a bit surprising that reviews of the lie detection training literature have been cautiously optimistic about our abilities to train lie detectors (Bull, 1989; DePaulo, 1994; Garrido & Masip, 1999). Garrido and Masip (1999) argued that given the lack of behavioral clues to lying, and police reliance upon stereotypes for their judgments, it would be very difficult to train police to be better lie catchers (although they suggest we should try). DePaulo (1994) was more optimistic, arguing that although the evidence is not solidly in favor of being able to train laypeople to detect lies, people know more about lie clues than they can typically express and thus we should be able to train not just laypeople but professionals.

Theoretical Rationale for Training

Ekman (1992) defined a lie as a deliberate attempt to mislead, without the prior consent of the target. Virtually all deception scholars agree on the deliberate attempt to mislead criterion of a lie (e.g., Knapp & Comadena, 1979; Levine, 1994; Miller & Stiff, 1993); however, there is not as much agreement upon the prior consent of the target criterion (e.g., Bok, 1978; Nyberg, 1993). Ekman (1992) suggests that certain social interactions imply that one will not be truthful, and that a person consents to that possible misinformation by engaging in that social interaction. For example, by entering a negotiation, one consents to the other not truthfully opening his or her bidding with his or her final asking price. Moreover, one consents to suspending reality and the true identities of the actors when attending a play or movie. The same applies to polite behavior, as when we show appreciation for a gift that we may not like. These latter lies are referred to as "white lies" to connote the social sanction of telling such lies. A person who speaks truthfully about every negative feeling or impression he has about gifts or people would become socially ostracized. We are raised not to attend to the real meaning behind these lies; in fact, it would be impolite to do so. Thus, we suggest that detecting these types of lies is not important to the average person, let alone

the professional lie catcher. In other words, it doesn't seem critical to society to train people to better catch these types of polite lies.

Yet there is a strong theoretical rationale for why these types of lies would be different from the lies faced by law enforcement, and this main difference is the presence of strong emotions in the high stake lies. Although a number of deception researchers have reasoned that lies are betrayed by both cognitive and emotional signals (e.g., deTurck & Miller, 1985; Ekman, 1992; Ekman & Frank, 1993; Feeley & deTurck, 1998; Greene, O'Hair, Cody, & Yen, 1985; Hocking & Leathers, 1980; Vrij, 2000; Zuckerman, DePaulo, & Rosenthal, 1981), it has been the emotion researchers who have most clearly elaborated the role of emotion in the detectability of lies. For example, Ekman's neurocultural theory of emotions argues that some basic emotions, such as anger, contempt, disgust, fear, happiness, sadness, or surprise are physiological reactions, derived by Darwinian evolutionary principles, that assist the organism to survive or thrive in recurrent life situations such as fleeing predators/enemies or fighting (Ekman, 1992). Part of the physiological reaction of a basic emotion is a facial expression that serves to signal an imminent behavior (such as fleeing in fear or attacking in anger). The nature of these basic emotions is that they are unbidden, and thus the impulse for the facial expression that accompanies that emotion is also unbidden (e.g., Rinn, 1984). However, Ekman argues that one's culture teaches one how to manage or control facial expressions; in other words, culture teaches its members facial "display rules" (Ekman, 1972). Yet when these basic emotions are elicited, and they are of sufficient intensity, the expressions will leak despite the efforts of the person to hide or mask them (Ekman, 1972; Ekman, O'Sullivan, Friesen, & Scherer, 1991). These facial expressions are recognized at rates greater than chance by all cultures (e.g., Ekman, 1994; Izard, 1994; see also Russell, 1994).

It is the presence of high stakes that is central to a liar feeling emotions when lying (Ekman, 1992). For example, the person interrogated by police would in all likelihood feel fear of getting caught, as getting caught means going to jail, which is a strong punishment. Thus the facial expression of fear can betray deception (Frank & Ekman, 1997). Note that high stakes can generate other reliable emotional clues to deceit beside fear. If the lie is personally relevant (e.g., about an action committed by the liar) then the liar is more likely to feel guilt or shame about the lie (e.g., Lazarus, 1991). Or, if there is excitement over getting away with the lie, duping delight, or happiness, can be elicited (Ekman, 1992).

In contrast, based on the neurocultural theory of emotions, one can predict that white lies are less likely to elicit emotion and there will not be as strong a fear of getting caught. It can also be predicted that there will be fewer feelings of guilt or distress over telling the white lie. A person lying about whether he or she enjoys a gift will not face jail time, and instead will be commended for being polite and showing tact. A person lying to a judge or police officer can be arrested or jailed. Thus, white lies and high stake lies should differ markedly in the type of information and signals generated by the liar. These signals can be useful to a lie catcher and ultimately useful for training lie catchers.

Besides emotions, other features of deception situations may serve to generate reliable clues to lying. For example, a skeptical target asking multiple questions will more strongly tax the cognitive system of the liar, producing "thinking" clues to deceit such as speech errors, reduced illustrators, and so forth (e.g., deTurck & Miller, 1985; Ekman, 1992; Feeley & deTurck, 1998; Greene, et al., 1985; Hocking

& Leathers, 1980; Vrij, 2000). We would expect to see some of these clues in both high and low stake lies, as both types can require cognitive effort to conceal one's true attitudes.

Not only do these stakes for accurate or inaccurate lie and truth judgments apply to individuals, but also to society. As described earlier, the harm-doers may be freed to continue terrorizing society, and the harmless may go to jail where they no longer contribute to society. Therefore, in contrast to the polite lies, it is important to be able to catch these high stake lies. Thus we believe that we should examine how well we can train people to better distinguish lies from truth. Based on the arguments above, however, the greatest impact for training should be found with attempting to train professional lie catchers to detect high stake lies.

Challenges to Lie Detection Training

There are many acceptable ways to train professional lie catchers, any one of which can be useful to inform us whether we can improve lie detection skills. However, based on the theoretical model described above, we feel that we can derive certain principles that can be developed from which to challenge previous training studies. Researchers can then examine not only whether these studies have provided a fair test as to whether people can be taught to improve their lie catching ability, but whether they over or under estimate the amount of the improvement. This article itemizes these into six challenges, and we will examine the extent to which published research on lie detection meets these challenges, in order to determine whether lie detection training can be worth its cost to society.

These six challenges must be met in order to allow us to judge whether training will benefit society. The first challenge is to create a deception situation that is relevant to professional lie catchers. Challenge 2 is to verify that there are behavioral clues to deceit in these training/testing materials. Challenge 3 is to create adequate training techniques executed over an appropriate period of time. Challenge 4 is to create adequate and appropriate pre and posttests of deception detection ability to provide a fair test and to monitor progress in training. Challenge 5 is to show that this training will generalize to new deception materials. Challenge 6 is to show that possible gains in training persist over time.

Challenge 1: Relevance

The first challenge for research in lie detection training is to create a deception situation that is relevant to what professionals face during interrogation. This is critical for a number of reasons. First, more relevant deception paradigms will maintain the interest of professionals because it will be clear to them why this is important to their careers—and motivation seems to be one factor associated with increased detection accuracy of professional lie catchers (Ekman & O'Sullivan, 1991; Ekman et. al., 1999). Second, the presence of stakes for the liar and truth teller increases the probability of detectable behavioral clues (Ekman & Frank, 1993), and we would predict that these clues would be more similar to the clues professionals see in their real world work.

The way to make the deception paradigm more relevant is to map out the value of the structural features of a deceptive situation faced by professionals and then try to recreate them in the laboratory (Frank, 1992; Podlesny & Raskin, 1977). The structural features of a deceptive situation refer to those invariants of all deceptive

situations and can be separated into the structural features of a lie and the lie situation. For example, each lie can include a number of features: (a) the type of lie (lie about an action, or opinion), (b) the form of the lie (concealment versus falsification), and (c) the motive behind the lie (to avoid punishment, or to protect another's feelings).

Each lie situation can also include: (a) characteristics of the liar, (b) characteristics of the target, (c) the relationship between the sender and the target, (d) a mode of inquiry, (e) the stakes associated with successful or unsuccessful deception, (f) the period of time between the decision to lie and the commission of the lie, and (g) the presence or absence of others. By no means would these be the only factors to consider, but they represent the more critical factors needed to approach ecological validity (Frank, 1992).

If we take a look at a hypothetical scenario commonly faced by professional lie catchers, such as a deceptive suspect in a police interview, the situation typically involves a fabrication lie about one's actions, and the suspect is typically motivated to avoid punishment. The interrogation also features a younger suspect and an older authority figure target who are not known to each other, and the suspect has typically 30 minutes or more to consider his or her lie. There is sometimes another investigator present, or a lawyer, but often not. The target is also able to ask multiple questions, and the successful lying suspect knows that he or she will get away with money, assault, or murder (depending upon the crime) and the unsuccessful lying suspect knows that he or she will be arrested and subject to trial (the stakes).

We will argue that a deception paradigm employed in any training study should share many of the structural features stated above in order to be relevant to a professional lie catcher, and to give any training effect the maximum opportunity to manifest itself. We of course acknowledge that there are ethical constraints in laboratory research that would make it impossible to recreate exactly some of the conditions found in the real world.

Challenge 2: Stakes

Once the structural features of a lie situation used to create the stimulus materials for training are similar to a real world situation faced by a professional lie catcher, then we must make sure this lie situation produces observable clues to lying. It does not seem prudent to train professionals on situations in which there are no clues that distinguish truths from lies, or in which we did not bother to measure whether there are clues that distinguish truths from lies. Because if there are no clues that distinguish lies from truths, then on what do we train professionals?

Although the previous literature can guide us in our search for trainable clues to deception/truthfulness, we believe that we cannot rely exclusively on the deception literature for these clues. This is because very few studies have employed high stake emotion-eliciting paradigms, and thus underrepresent potential emotional clues to deceit. For example, two studies using high stakes have been able to distinguish lies and truths at rates greater than 80%, based mainly on facial expressions of emotion (Ekman et al., 1991; Frank & Ekman, 1997). Therefore, training on a high stake paradigm is not only relevant, but also crucial to any successful training. Moreover, facial expressions and other evidence of emotion

have universal, cross-cultural interpretation (e.g., Ekman, 1994; Izard, 1994) and have similar meanings across different lie situations—thus rendering them even more useful for training (Frank & Ekman, 1997). Therefore, we argue strongly that any deception paradigm used to train professional lie catchers must have high stakes, not only to be relevant, but to generate more behavioral clues—clues that should generalize across high stake situations and time.

Challenge 3: Training

Once we have relevant and reliable deception material, we must create training techniques that will transmit this knowledge effectively. All reputable deception researchers agree that detecting deception from demeanor is complex (Anderson, DePaulo, Ansfield, Tickle & Green, 1998; Buller & Burgoon, 1996; DePaulo et al., 1985; Ekman & Frank, 1993; Feeley & Young, 1998; Kraut & Poe, 1980; Levine, Park, & McCornack, 1999; Zuckerman et al., 1981). The big question is how does one transmit this complex knowledge to time-limited professional lie catchers? It seems that one might underestimate any training effect by simply giving feedback as to the correct or incorrect responses, without actually pointing out the variables that are responsible for that judgment. Moreover, given the complexity of the relationships among these variables, it seems that formal training, in conjunction with feedback, should be for at least the equivalent of one classroom period (50 minutes to an hour), and ideally repeated over a number of days for maximum retention. We do not expect our students to become fluent in the material we present in our lectures immediately after we present it; we expect them to go home, read supplemental materials, and give them weeks to memorize and comprehend before testing them. Training a skill such as deception detection takes practice.

We believe professional lie detectors should receive well-presented information, ample opportunity to practice, and be given feedback as to their performance. Anything less, such as only giving feedback (e.g., Zuckerman, Koestner, & Colella, 1985), or only training for a few minutes (e.g., Vrij & Graham, 1997), may seriously underestimate whether one can train professionals to be better detectors of deceit.

Challenge 4: Testing

The next challenge is to insure that there are adequate pre and posttest measures of training. What this means is that professionals should be assessed on their skill prior to engaging in the training, and then receive training (treatment) and others not (control) in a basic pretest-posttest design (Carlsmith, Ellsworth, & Aronson, 1976). Although this seems like a simple recommendation, and there are other techniques equally valid, such as random assignment to conditions in posttest only designs, there are a number of potential pitfalls that complicate this step. First, not only should the stimuli persons differ within the pre and posttest, but also they should differ between the pre and posttest. Otherwise, this test might introduce deception detection strategies based on heuristics having to do with the design of the stimulus test, rather than the behavior shown in the stimulus test. For example, if a lie catcher sees a stimulus person in the posttest whom they remember from the pretest, they may judge this person the same in the posttest as the pretest due to the residual impression of that person created by the pretest. Or,

they may gainsay their pretest judgment by trying to outguess the experimenter—reckoning that if the pretest exposure was a lie, then the posttest must be a truth (or vice versa). Either way, this suggests that the form of the testing and assessment introduces artificial decision rules specific to that testing and assessment rather than decision rules that will generalize to the real world. There are data consistent with this reasoning; for example, with repeated viewings, lie catchers view the first instance of a stimulus person as a baseline, and the second as a deviation from that baseline and hence subjects are more likely to judge the second instance as deception, independent of actual truth or deceptiveness (O'Sullivan, Ekman, & Friesen, 1988). This is not to argue that multiple viewings of the same stimulus person are not ecologically valid; they are. A police interview will involve many topics, and may feature multiple lies and truths uttered by the same suspect over the course of the interview (e.g., Vrij, 2000). However, we suggest researchers attend closely to the composition of the pre and posttests to reduce the likelihood of any of this artificial heuristic decision-making.

Second, professionals should be given a vague expectation about the number of liars or truthtellers in the tests. Our experience in lie detection research has been that on more than one occasion we had a law enforcement officer judge all stimulus persons as liars, because of the presumption that everyone lies all the time—a simple heuristic that they would not employ outside the job. Moreover, we have had students judge all stimulus persons as truthful, and probably for a variety of reasons. Researchers have also documented a truth bias in people's judgments in lie situations (Feeley, deTurk, & Young, 1995; Levine & McCornack, 1992; Levine, Park, & McCornack, 1999; McCornack & Parks, 1986). To avoid these potentially confounding rationales, we find it helpful to tell our lie catchers that at least one-fourth of the stimulus persons are truthful, and one-fourth deceptive. This information is specific enough to let them know that the pre or posttest is not a trick, and yet vague enough that they do not engage in another heuristic lie detection strategy based on our revealing the exact proportion of liars and truthtellers in either the pre or posttest. For example, if lie catchers are told exactly half the stimulus persons on the video are lying, and lie catchers are then presented with a ten item test, anecdotal evidence from our labs suggests that once they make five judgments of truth (or lie), they judge the remaining items as lies (or truths) so that they arrive at exactly five truth and five lie judgments. Thus, they are making judgments based on another artificial heuristic driven by the composition of the test, rather than the behavior seen in the video. We acknowledge that giving subjects a hint may also constrain lie catchers' judgments, but certainly not to the degree that giving them the exact proportion does. We also recommend having an equal number of liars and truthtellers to enable one to assess easily whether scores are above or below chance. Levine, Park, and McCornack (1999) recently documented a veracity effect in lie detection whereby students tend to achieve significantly greater accuracy when judging truth tellers versus judging liars (see also Ekman et al., 1999; Feeley et al., 1995). However, Levine et al. (1999) reported student lie detection accuracy rates consistently near or below chance level. Levine et al.'s (1999) findings indicate that deception detection accuracy is both a function of the truth or lie bias of the targets and the ratio of truths and lies being judged. Thus, in the case of the lie-biased police officer, accuracy rates will certainly differ as a function of the number of truthful and deceptive senders presented in the stimulus materials. Although techniques exist for controlling for

guessing in tests with unequal numbers of liars and truth tellers (see Wagner, 1993), we still feel the tests should be balanced to more easily control for guessing.

Third, the pre and posttest should feature adequate sample sizes. One cannot have a good training program based on uncovering the lies of a single stimulus person, given that people vary wildly in their abilities to lie successfully (Bond, Kahler, & Paolicelli, 1985; Zuckerman, DeFrank, Hall, Larrance, & Rosenthal, 1979). If this stimulus person were too easy or *leaky*, we would get an inflated view of training effects. By contrast, if this stimulus person were particularly adroit at masking his or her lies, we would underestimate training effects. For example, Vrij and Mann (2001) provided useful insight into detecting real world lies of a murderer; however, their results were limited by the fact that they looked at only one murderer, who may or may not be a convincing liar. We suggest ten as a minimum sample size, given that ten is the number minimally expected in any ANOVA cell (Snedecor & Cochrane, 1980). More stimulus persons will of course be better and more reliable—but within a limit. Our experience also tells us that if the judgment task takes more than 30 minutes, lie catchers, particularly college sophomores, become tired, distracted, and stop concentrating on their judgments. Therefore, the fourth challenge is to design pre and posttests that allow adequate internal and external validity, without causing lie catchers to revert to artificial heuristics for their judgments.

Challenge 5: Situational Generality

It is one thing to train lie catchers on a circumscribed batch of stimulus material that one can thoroughly and exhaustively study and mine for verifiable behavioral clues, but it is quite another to see if the training based upon these clues generalizes to different high stake deception situations. This is a big reason why we issued challenges about using a high stake paradigm, with at least 10 stimulus persons per pre or posttest when training lie catchers. In fact, Frank and Ekman (1997) found that undergraduates' lie catching abilities did generalize across different high stake lies. This bodes well for training across situations.

However, the main issue here is to show that this training is not specific to the narrower training materials, thus insuring external validity of the training. After all, the intended social impact of this research is to help professionals process accurately the lies and truths of the real world, and our training, if effective, should be able to meet this challenge. Of course, this implies that there are highly reliable lying clues and research has not found this to be the case (DePaulo et al., 1985; Ekman, 1992; Zuckerman, et al., 1981). However, clues that are probabilistically related at 80% accuracy or higher would be immensely useful toward raising accuracy rates beyond the 55% of the average lie catcher (DePaulo et al., 1985; Feeley & Young, 1998; Kalbfleisch, 1992; Kraut, 1980). Accuracy rates as high as 80% would make lie catchers almost as accurate as a polygraph or lie detector machine (Saxe, Dougherty, & Cross, 1985). Therefore, the fifth challenge to training effects is to show that they will generalize to different high stake lie situations, and ultimately to the real world.

Challenge 6: Time Generality

The last challenge is to insure that this training has an impact that exceeds the posttest. Ideally, a true training effect should show that professional lie catchers

should outperform not only a comparable control group, but they should outperform their own baseline levels of accuracy one week, one month, or even one year later. There are many factors that work against the persistence of training effects over time. For example, stereotypes about deceptive behavior may be so strong that they reemerge with little provocation (e.g., Anderson et al., 1998; DePaulo, 1994; Feeley & Young, 2000; Zuckerman, Koestner, & Colella, 1984) and may need to be fought constantly with repeated training over the life of the professional lie catcher.

These are the challenges facing the question of whether training effects exist for professional lie catchers. The suggested procedures allow training effects the greatest chance of emerging, while limiting the possibility of a Type 1 error. We note that studies that fail to employ these steps are not necessarily flawed, however, but may be increasing the possibility of a Type II error, or at least underestimate training effects.

A Quantitative Review of Current Research in Lie Detection Training

There is a growing body of research that has sought to test the notion of lie detection training. A meta-analysis was performed to examine the aggregate effect size for training compared to the control group (i.e., no training). We used several search procedures in an effort to include all training studies relevant to lie detection training. First, an ancestry search was done by closely examining the reference sections of all training studies known to the two authors and recent books on the subject (e.g., Vrij, 2000). Next we contacted several experts in lie detection training by telephone and internet for any possible unpublished or difficult to locate studies. We then did several electronic searches of academic databases including FIRST SEARCH, ERIC, PsycINFO and Expanded Academic ASAP using several combinations of search terms including "training," "lying," "lie detection," "deception," "deception detection," and "learning."

We considered the study a training study if lie detectors in the treatment condition were given additional information related to "how to catch a liar" (e.g., deTurck et al., 1997; Vrij & Graham, 1997) or if lie detectors were given feedback regarding their success (or lack of success) on earlier trials during or before the lie detection task (e.g., deTurck & Miller, 1990; Fiedler & Walka, 1993; Zuckerman et al., 1984). Additionally, all training conditions had to be compared to a control condition that did not provide additional information and/or feedback.

Our search yielded 20 relevant comparisons across 11 published studies on lie detection training. Table 1 outlines the size of the sample examined, the effect size for each comparison, and the mean accuracy rates for training and the control condition. Very few studies reported measures of variation (e.g., standard deviations) and many studies neglected to report overall means (or means for each treatment condition). Formulas outlined by Rosenthal (1994) were used to convert reported statistics to a common effect size metric (the r statistic). Most effect sizes in published studies were reported in t-tests, F-tests or Pearson correlations, which are easily converted to r-values (e.g., Cooper, 1984; Rosenthal, 1994). Table 1 reports the effect size for each training condition when compared to the control condition.

A fixed effects meta-analysis (see Hedges, 1994) was conducted using a weighted effect size for each study. The number of participants across the 20

TABLE 1
Summary of Means, Power, and Effect Sizes for Training and Control Groups

Study	n	r^p	Overall		Truth		Lie	
			T	C	T	C	T	C
deTurck (1991)*	183	−.33	.69	.55	—	—		
deTurck et al. (1997)								
Visual	82	+.18	.62	.57	.51	.56	.73	.59
Vocal	82	+.00	.55	.57	.45	.56	.65	.59
Both	83	−.18	.64	.57	.56	.56	.73	.59
deTurck et al. (1990)*	188	−.26	.65	.54	—	—	—	—
deTurck & Miller (1991)*	390	−.08	.61	.53	—	—	—	—
Fiedler & Walka (1993)								
Inform	48	+.56	.65	.56	.62	.56	.68	.56
Inform/Feedback	48	+.59	.65	.56	.62	.56	.68	.56
Kassin & Fong (1999)	40	.33	.46	.55	—	—	—	—
Kohnken (1987)								
Nonverbal	40	−.10	.42	.47	.53	.63	.32	.32
Content	40	−.03	.48	.47	.69	.63	.26	.32
Speech	40	−.13	.40	.47	.48	.63	.33	.32
Vrij (1994)								
Information	240	−.13	.51	.49	—	—		
Information/Feedback	240	+.13	.53	.49	—	—	—	—
Vrij & Graham (1997)*	69	+.28	.51	.48	—	—	—	—
Zuckerman et al. (1984)								
Four After	65	+.00	.61	.62	—	—	—	—
Eight After	64	+.36	.70	.62	—	—	—	—
Four Before	105	+.19	.66	.62	—	—		
Mixed	67	+.17	.70	.62	—	—		
Zuckerman et al. (1985)*	117	+.36	.62	.58	—	—		
Means		+.20[b]	.58[c]	.54[c]	.56[c]	.58[c]	.55[c]	.49[c]

Note. Asterisks represent studies without main effects means reported for training. Means were taken from tables or deduced from interaction means reported, thus some means may be inexact. T = Training Condition; C = Control Condition. Control means were used only once in unweighted means. Superscript *a* indicates a weighted effect size, often this is denoted by T_ (Cooper & Hedges, 1994). The weighted effect size combines the raw effect with power of study. Superscript *b* indicates a weighted combined effect size. Superscript *c* indicates that all means reported are unweighted (i.e., these means do not take into consideration the size of the sample).

comparisons was 2,233 with 1,072 persons in the training conditions and 1,161 in the control conditions. Across all 20 cases the fixed-effects weighted mean *r* was .20 and the 95% confidence interval for this mean was .16 to .24, indicating a dependably positive overall effect for training. The test of homogeneity yielded an significant value, $Q(19) = 79.11$, $p < .001$. Thus, the variance is much larger than what would be expected by chance alone. We caution readers against reading too much into the mean *r* value of 0.20 considering that 16 of the 20 effect sizes reported in Table 1 are outside of the 95% confidence interval and the *r* values range from −.33 to +.59.

Our interpretation of these meta-analytic results suggests two conclusions. First, there is a small but dependable gain in detection accuracy for training across 20 paired comparisons in 11 published studies in lie detection training. Table 1 outlines the detection accuracy means for these 11 studies and finds a 4% gain in accuracy for trained observers. It should be noted that this is an unweighted mean difference. Second, we suggest that the gains documented in lie detection training research in actuality may underestimate the possible gains of detection training if

current research methodology met or even came close to the challenges set forth in this paper for research in training. Moreover, while the research synthesis found a statistically significant and dependable effect when comparing training accuracy means to control means, we still question the practical significance of a .58 mean accuracy rate for training. As mentioned, many of the studies included in the meta-analysis do not meet the challenges raised earlier in this paper. With this said, we feel it premature to answer the question, "Can we train observers to catch liars?"

Does Current Research Measure Up?

The meta-analytic findings also suggest that there may be components of the research designs and stimulus materials that account for an appreciable amount of the variance in training. Perhaps if future research met more or all of the above-mentioned challenges we could gain a more reliable estimate of the pure effects of training.

The Challenge of Relevance

The first challenge to training research is to emulate a deception detection situation that would typify what a professional would confront. The suspect should face (and be aware of) palpable consequences for being labeled a liar or a truth teller. Other structural aspects of the lie must also be relevant to a forensic setting. For example, the liar would most likely be fabricating or omitting information about his or her actions (or witnessing another's actions) and the interrogator should be somewhat skeptical of the source's veracity.

Current research does fairly well in meeting some of the aspects of the relevance challenge. Consider the dot-estimation paradigm used by deTurck and colleagues (deTurck et al., 1997; 1990). Students in these studies (see also Dulaney, 1982; Exline, 1971; Feeley & deTurck, 1995; Stiff & Miller, 1986) were negatively probed about the strategies they and a partner (confederate) reported using to count a seemingly infinitesimal number of dots on a set of cards. Some of the students, however, were coaxed into cheating on the task by a partner-confederate who opened the experimenter's folder that had the answers to the number of dots on a sheet of paper. Thus any reported strategies other than cheating would be deceptive on the part of the student. This dot-estimation paradigm is somewhat similar to the lie detection task of the professional. The behavior under question is relevant (i.e., about your own actions or ideas) and there is a skeptical target (an experimenter) interrogating the subject. However, there are not clear incentives or punishments for successful or unsuccessful deception. There is a reward offered for the closest dot-estimations that may be a carrot for the successful lie. One might also argue that getting caught while cheating on a lab task by an authority figure (i.e., a professor) may be a worrisome to a naïve research participant. A recent study by Feeley and deTurck (1998) found only one significant difference in behavior (speech length) between those who cheated on the dot-estimation task and those who did not cheat on the task which would suggest that the incentives may not be as motivating as once thought (e.g., deTurck & Miller, 1985).

Three studies (Kassin & Fong, 1999; Vrij, 1994; Vrij & Graham, 1997) attempted to raise the stakes for liars by involving students in a mock crime operation in the laboratory. Vrij (1994; see also Vrij & Graham, 1997) had students participate in

two interrogations by uniformed police officers. Students were instructed to deny the accusations of the officers that they possessed a pair of headphones—in one of the two interviews each student did have the headphones in his or her possession and was thus lying in one interview and telling the truth in the other interview. Kassin and Fong (1999) had students caught in the act of committing a mock crime. Students vandalized a campus wall, stole jewelry (or a stuffed animal), trespassed, or broke into a computer laboratory. Students in the Kassin study were then immediately interrogated by an incensed, 48-year-old police detective who read students their Miranda rights before questioning commenced.

Taken together, most of the 11 experiments pass muster for the initial criterion of challenge 1—the personal relevance of the lie to the participant and the real world. Moreover, in many of the studies (e.g., deTurck et al., 1997; Kassin & Fong, 1999), participants were often questioned by a skeptical target. Four experiments failed to employ relevant lies in their experimental design. Zuckerman et al. (1984) had students truthfully or untruthfully describe personal acquaintances and participants in two of the deTurck (deTurck, 1991; deTurck & Miller, 1990) experiments were instructed to describe pleasant/unpleasant slides truthfully or untruthfully. Finally Druckman, Rozelle, and Baxter (1982) used an actor who played the role of a Soviet ambassador. The actor recited lines from a script and required several takes to send his message.

Meeting the Challenge of High Stakes

A plausible explanation for the notoriously low deception detection accuracy rates reported in laboratory studies is that liars and truth tellers simply do not behave differently. Most studies in lie detection have examined lies of little, if any, consequence and few studies bother to verify if there are any behavioral differences between liars and truth tellers (for exceptions see Vrij, 1994; Fiedler & Walka, 1993). Thus, the second challenge requires that future research in training use lies and truths with higher stakes. Also, there should be significant differences in body, face, voice, or speech to verify that the stakes for success are acceptable.

There are two studies that examine (and find) behavioral differences between liars and truth tellers. Vrij (1994) found liars to make fewer hand and finger movements during deception. Fiedler and Walka (1993) found veracity to be a function of seven nonverbal cues (smiling, hand movements, adaptors, pitch, speech hesitations, pauses, channel discrepancies) that were identified as reliable clues to lying by Zuckerman and Driver (1985). The other nine studies failed to examine for behavioral differences between liars and truthtellers. We suspect if they did, there would not be significant findings for two reasons. First, the lies told were white lies (e.g., Zuckerman et al., 1984) or lies told to appease the instructions of the experimenter (e.g., Kohnken, 1987). Second, there are so few senders in many of the studies that there would be not be enough in a cell to perform parametric statistics.

The existing research on lie detection, on the whole, fails to meet the challenge of high stakes. To adequately assess our ability to train lie detectors it is imperative that lie detectors have something to go on when viewing liars and truthtellers. Lies that are worth catching are typically lies that stir up greater emotion that, in turn, disrupt cognition. Emotion and cognitive disruption typically leave identifiable clues for the receiver to pick up on.

Meeting the Challenge of Training

Challenge 3 recommends that training of lie detectors include two minimal criteria: length and feedback. That is, training sessions should be no less than 60 minutes to one hour and lie detectors should be given the opportunity to practice and/or given feedback regarding their performance. It should be noted that we set 60 minutes of training as an absolute minimal criterion. To properly train there should be multiple training sessions and students of lie detection should also be given supplemental reading materials. Only two studies (Kassin & Fong, 1999; Kohnken, 1987) trained for 1 hour or more and no study examined training with multiple instructional sessions. With respect to feedback, both Vrij (1994, information plus feedback condition) and deTurck and colleagues (deTurck, 1991; deTurck et al., 1997; 1990; deTurck & Miller, 1990) provided feedback prior to judging stimulus persons. It is worth noting that deTurck's experiments have yielded relatively high training accuracy rates generally in the range of 60–70% accuracy. Vrij found higher judgmental accuracy with those training combined with feedback when compared to training alone or the control group. Zuckerman et al. (1984) and Fiedler and Walka (1993) provided feedback during the lie detection task. Results from these studies indicate mixed results. Fiedler and Walka found feedback to increase truth accuracy but not lie accuracy and Zuckerman et al. found feedback to help within senders but not between senders. Stated differently, lie detection improved as function of learning for some senders but this learning did not help receivers judge other senders. This would suggest that lie detectors were picking up on person-specific clues that did not generalize to other senders.

Meeting the Challenge of Testing

The fourth challenge has to do with the experimental methods used to compare trained observers to untrained observers. On this challenge there are two main criteria. First, research has used very few stimulus persons—that is, lie detectors are only detecting deception of 4 or 8 senders on the videotape. We recommend a minimum of 10 senders as a minimum. Thus, only Fiedler and Walka (1993), Vrij (1994), and Vrij and Graham (1997) include an adequate number of stimulus persons in their training experiment.

Experiments in training should also employ true experimental pretest/posttest designs (Snedecor & Cochrane, 1980). Participants should be randomly assigned to one of two conditions: control group (i.e., lie detection without training) and a treatment group (training manipulation). Participants in each group should judge different stimulus persons between the pretest and posttest. To date, there are no studies that have used this design. Most studies employ a one-shot design comparing trained observers' accuracy to untrained observers' accuracy which is acceptable if randomization is insured (few studies discuss how participants were assigned to conditions) but provides less control than the pre and posttest design.

Meeting the Challenge of Generalizability

None of the studies examined the generalizability of the clues used for training or the effects of training over time. For training to be successful one must be able to detect different types of lies and also be able to detect deception in different

stimulus persons over an extended period of time, not just immediately after training. Some studies used cues reported to be associated with deception in previous research (e.g., deTurck & Miller, 1990; Fiedler & Walka, 1993) and applied these cues to different stimulus persons (and different lies). Zuckerman et al. (1984) indirectly tested the generalizability of person-specific feedback and found that learning effects did not generalize across different persons.

No study has examined the longitudinal effects of training. We would recommend testing judgmental accuracy over time. If judgmental accuracy does increase as function of behavioral cue training, do these effects persist over time? Is training adequate and sufficient to invoke longitudinal effects?

Summary and Future Directions

Taken together, the current research does not meet many of the challenges we have set down. First, some of the deception situations would be irrelevant to the professional lie catcher and fewer still confirmed the presence of behavioral clues that would distinguish truths from lies. Most did not use adequate training techniques (time length, stakes of lie, clues) and none have measured directly the generalizability of training effects over situation and time. Yet, many of the studies still show a slight training effect. We will argue that these studies are not flawed; in fact, it appears as if they were designed to see whether it is even possible to train. In this, they accomplished their task, but in the process probably underestimate a training effect.

There are a great number of other issues to explore with training. For example, what does training improve? Does it improve lie catching, but not truth recognition? Does training improve for stimulus subjects showing facial clues, vocal clues, or body clues, or some combination of these clues? Moreover, we can ask to what extent does this training improve the ability of a lie catcher in a face to face situation, compared to the effects of training on a live observer or while viewing video. These issues are beyond the scope of this paper.

We feel that we have not yet fairly examined the extent to which one can train professionals to improve their abilities to distinguish truths from lies. We have employed techniques that overestimate our abilities to detect deception (e.g., using limited stimulus subjects) and techniques that underestimate our abilities (e.g., limited training time). Until we take the steps necessary, we believe that we should not sell short our ability to train effectively professional lie catchers at rates that will meaningfully affect the safety and well being of individuals in our society.

References

Anderson, D. E., DePaulo, B. M., Ansfield, M. E., Tickle, J. J., & Green, E. (1998). Beliefs about cues to deception: Mindless stereotypes or untapped wisdom? *Journal of Nonverbal Behavior, 23*, 67–89.

Bok. S. (1978). *Lying: Moral choice in public and private life.* New York: Random House.

Bond, C. F., Kahler, K. N., & Paolicelli, L. M. (1985). The miscommunication of deception: An adaptive perspective. *Journal of Experimental Social Psychology, 21*, 331–345.

Bull, R. (1989). Can training enhance the detection of deception? In J. C. Yuille (Ed.), *Credibility assessment* (pp. 83–99). Norwell, MA: Kluwer Academic Publishers.

Buller, D. B. & Burgoon, J. K. (1996). Another look at information management: A rejoinder to McCornack, Levine, Morrison, and Lapinski. *Communication Monographs. 63*, 92 -98.

Carlsmith, J. M., Ellsworth, P. C., & Aronson, E. (1976). *Methods of research in social psychology.* New York: Random House.

Cooper, H. M. (1984). *Integrating research: A guide for literature reviews* (2nd ed.). Newbury Park, CA: Sage.

Cooper, H. M., & Hedges, L. V. (1994). *The handbook of research synthesis*. New York: Russell Sage Foundation.

DePaulo, B. M. (1994). Spotting lies: Can humans learn to do better? *Current Directions in Psychological Science, 3*, 83–86.

DePaulo, B. M., & Pfeifer, R. L. (1986). On-the-job experience and skill at detecting deception. *Journal of Applied Social Psychology, 16*, 249–267.

DePaulo, B. M., Stone, J., & Lassiter, D. (1985). Deceiving and detecting deceit. In B. R. Schlenker (Ed.), *The self and social life* (pp. 323–370). New York: McGraw-Hill.

deTurck, M. A. (1991). Training observers to detect spontaneous deception: The effects of gender. *Communication Reports, 4*, 79–89.

deTurck, M. A., Feeley, T. H., & Roman, L. (1997). Vocal and visual cue training in behavioral lie detection. *Communication Research Reports, 14*, 249–259.

deTurck, M. A., Harszlak, J. J., Bodhorn, D., & Texter, L. (1990). The effects of training social perceivers to detect deception from behavioral cues. *Communication Quarterly, 38*, 1–11.

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.

Druckman, D., Rozelle, R. M., & Baxter, J. C. (1982). *Nonverbal communication*. Beverly Hills: Sage.

Dulaney, E. F. (1982). Changes in language behavior as a function of veracity. *Human Communication Research, 9*, 75–82.

Dwyer, J., Neufeld, P., & Scheck, B. (2000). *Actual innocence: Five days to execution and other dispatches from the wrongly convicted*. New York: Doubleday.

Ekman, P. (1972). Universal and cultural differences in facial expressions of emotion. In J. K. Cole (Ed.), *Nebraska symposium on motivation* (pp. 207–283). Lincoln: University of Nebraska Press.

Ekman, P. (1992). *Telling lies: Clues to deceit in the marketplace, politics, and marriage*. New York: Norton.

Ekman, P. (1994). Strong evidence for universals in facial expression: A reply to Russell's mistaken critique. *Psychological Bulletin, 115*, 268–287.

Ekman, P., & Frank, M. G. (1993). Lies that fail. In M. Lewis & C. Saarni (Eds.), *Lying and deception in everyday life* (pp. 184–200). New York: Guilford Press.

Ekman, P., & O'Sullivan, M. (1991). Who can catch a liar? *American Psychologist, 46*, 913–920.

Ekman, P., O'Sullivan, M., & Frank, M. G. (1999). A few can catch a liar. *Psychological Science, 10*, 263–266.

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

Exline, R. V. (1971). Visual interaction: The glances of power and preference. *Nebraska Symposium on Motivation*, 163–206.

Feeley, T. H., & deTurck, M. A. (1995). Global cue usage in behavioral lie detection. *Communication Quarterly, 43*, 420–430.

Feeley, T. H., & deTurck, M. A. (1998). The behavioral correlates of sanctioned and unsanctioned deceptive communication. *Journal of Nonverbal Behavior, 22*, 189–204.

Feeley, T. H., deTurck, M. A., & Young, M. J. (1995). Baseline familiarity in lie detection. *Communication Research Reports, 12*, 160–169.

Feeley, T. H., & Young, M. J. (1998). Humans as lie detectors: Some more second thoughts. *Communication Quarterly, 46*, 109–126.

Feeley, T. H., & Young, M. J. (2000). Self-reported cues about deceptive and truthful communication: The effects of cognitive capacity and communicator veracity. *Communication Quarterly, 48*, 101–119.

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. (1992). Commentary: On the structure of lies and deception situations. In S. Ceci, M. Leichtman, & M. B. Putnick (Eds.), *Cognitive and social factors in early deception* (pp. 127–146). Hillsdale, NJ: Erlbaum.

Frank, M. G., & Ekman, P. (1997). The ability to detect deceit generalizes across different types of high stake lies. *Journal of Personality and Social Psychology, 72*, 1429–1439.

Garrido, E., & Masip, J. (1999). How good are police officers at spotting lies? *Forensic Update, 58,* 14–21.

Greene, J. O., O'Hair, H. D., Cody, M. J., & Yen, C. (1985). Planning and control of behavior during deception. *Human Communication Research, 11,* 335–364.

Hedges, L. V. (1994). Fixed effects models. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 285–300). New York: The Russell Sage Foundation.

Hocking, J. E., & Leathers, D. G. (1980). Nonverbal indicators of deception: A new theoretical perspective. *Communication Monographs, 47,* 119–131.

Izard, C. E. (1994). Innate and universal facial expressions: Evidence from developmental and cross-cultural research. *Psychological Bulletin, 115,* 288–299.

Kalbfleisch, P. J. (1992). Deceit, distrust and the social milieu: Application of deception research in a troubled world. *Journal of Applied Communication Research, 20,* 308–334.

Kassin, S. M; & Fong, C. T. (1999). "I'm innocent!": Effects of training on judgments of truth and deception in the interrogation room. *Law & Human Behavior, 23,* 499–516.

Kohnken, G. (1987). Training police officers to detect deceptive eyewitness statements: Does it work? *Social Behaviour, 2,* 1–17.

Kraut, R. E. (1980). Humans as lie detectors: Some second thoughts. *Journal of Communication, 30,* 209–216.

Kraut, R. E., & Poe, D. (1980). Behavioral roots of person perception: The deception judgments of customs inspectors and laymen. *Journal of Personality and Social Psychology, 39,* 784–798.

Knapp, M. L., & Comadena, M. E. (1979). Telling it like it isn't: A review of theory and research on deceptive communications. *Human Communication Research, 5,* 270–285.

Lazarus, R. S. (1991). *Emotion and adaptation.* New York: Oxford University Press.

Levine, T. R. (1994, November). *Conceptualizing interpersonal and relational deception.* Paper presented at the Speech Communication Association, New Orleans, LA.

Levine, T. R., & McCornack, S. A. (1992). Linking love and lies: A formal test of the McCornack and Parks model of deception detection. *Journal of Social and Personal Relationships, 9,* 143–154.

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., & Parks, M. (1986). Deception detection and the other side of trust. In M. L. McLaughlin (Ed.), *Communication yearbook 9* (pp. 377–389). Beverly Hills, CA: Sage.

Miller, G. R. & Stiff, J. B. (1993). *Deceptive communication.* Newbury Park, CA: Sage.

Nyberg, D. (1993). *The varnished truth.* Chicago: The University of Chicago Press.

O'Sullivan, M., Ekman, P., & Friesen, W. V. (1988). The effect of comparisons on detecting deceit. *Journal of Nonverbal Behavior, 12,* 203–215.

Podlesny, J. A., & Raskin, D. (1977). Physiological measures and the detection of deception. *Psychological Bulletin, 84,* 782–799.

Rinn, W. E. (1984). The neuropsychology of facial expression: A review of the neurological and psychological mechanisms for producing facial expression. *Psychological Bulletin, 95,* 52–77.

Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper and L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 231–244). New York: Russell Sage Foundation.

Russell, J. A. (1994). Is there universal recognition of emotion from facial expressions? A review of the cross-cultural studies. *Psychological Bulletin, 115,* 102–141.

Saxe, L., Dougherty, D., & Cross, T. (1985). The validity of polygraph testing: Scientific analysis and public controversy. *American Psychologist, 40,* 355–366.

Snedecor, G. W., & Cochrane, W. C. (1980). *Statistical methods.* Ames, IA: The Iowa State University Press.

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.* New York: Wiley.

Vrij, A., & Graham, S. (1997). Individual differences between liars and the ability to detect lies. *Expert Evidence: The International Digest of Human Behaviour Science and Law, 5,* 144–148.

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

Wagner, H. L. (1993). On measuring performance in category judgment studies of nonverbal behavior. *Journal of Nonverbal Behavior, 17,* 3–28.

Zuckerman, M, DeFrank, R. S., Hall, J. A., Larrance, D. T., & Rosenthal, R. (1979). Facial and vocal cues of deception and honesty. *Journal of Experimental Social Psychology, 15,* 378–396.

Zuckerman, M., DePaulo, B. M., & Rosenthal, R. (1981). Verbal and nonverbal communication of deception. In L. Berkowitz (Ed.), *Advances in experimental social psychology, Volume 14* (pp. 1–59). San Diego: Academic Press.

Zuckerman, M. & Driver, R. E. (1985). Telling lies: Verbal and nonverbal correlates of deception. In W. A. Siegman & S. Feldstein (Eds.), *Multichannel integration of nonverbal behavior* (pp. 129–147). Hillsdale, NJ: Erlbaum.

Zuckerman, M., Koestner, R. E., & Alton, A. (1984). Learning to detect deception. *Journal of Personality and Social Psychology, 46,* 519–528.

Zuckerman, M., Koestner, R. E., & Colella, M. J. (1985). Learning to detect deception from three different communication channels. *Journal of Nonverbal Behavior, 9,* 188–194.