

## “He’s guilty!”: Investigator Bias in Judgments of Truth and Deception<sup>1</sup>

Christian A. Meissner<sup>2,4</sup> and Saul M. Kassin<sup>3,4</sup>

---

*Detecting deception is an inherently difficult task, but one that plays a critical role for law enforcement investigators in the interrogation room. In general, research has failed to indicate that performance in this domain is improved by training or prior experience. A signal detection framework is applied to the paradigm to better conceptualize the influence of these two factors. We found that although neither factor influenced discrimination accuracy, there was an effect on response bias such that training and prior experience appeared to increase the likelihood of responding “deceit” as opposed to “truth.” This “investigator bias” was observed both in a review of the literature and in this study of North American law enforcement investigators who took part in a forensically based deception-detection task. Possible theoretical mechanisms and practical implications of these findings are discussed.*

---

**KEY WORDS:** deception; interrogation; response bias.

Police-induced confessions are a potent prosecutorial weapon that can have far-reaching and rippling effects on the disposition of cases and on the criminal justice system as a whole. Indeed, modern police interrogations are so powerful that they have, at times, elicited coerced-compliant and -internalized false confessions from innocent people (Gudjonsson, 1992; Kassin, 1997; Leo & Ofshe, 1998; Radelet, Bedau, & Putnam, 1992; Scheck, Neufeld, & Dwyer, 2000). Research suggests that the process of interrogation is persuasive, if not too persuasive, in part because it is explicitly based upon a presumption of guilt—an assumption that itself can set in motion a number of cognitive and behavioral confirmation biases (Kassin, Goldstein, & Savitsky, 2001). As described by Inbau, Reid, Buckley, and Jayne (2001)

<sup>1</sup>Portions of this paper were presented at the 2002 American Psychology–Law Society Conference in Austin, TX.

<sup>2</sup>Department of Psychology, Florida International University, Miami, Florida.

<sup>3</sup>Department of Psychology, Williams College, Williamstown, Massachusetts.

<sup>4</sup>To whom correspondence should be addressed at Department of Psychology, Florida International University, University Park, Miami, Florida 33199; e-mail: meissner@fiu.edu or Department of Psychology, Williams College, Williamstown, Massachusetts 01267; e-mail: skassin@williams.edu.

in their manual, *Criminal Interrogation and Confessions*, "An interrogation is conducted only when the investigator is reasonably certain of the suspect's guilt" (p. 8).

Justifying the use of heavy-handed tactics, many law enforcement professionals believe that they can determine whether a suspect is being truthful or deceptive on the basis of a preinterrogation interview during which a suspect is questioned in a nonconfrontational manner about his or her knowledge, whereabouts, and possible involvement in the crime. Furthermore, many believe that they can learn to make these initial judgments at high levels of accuracy by observing various aspects of the suspect's verbal and nonverbal behavior (Inbau et al., 2001; Vessel, 1998). According to John E. Reid and Associates, for example, investigators trained in their Behavior Analysis Interview are able to distinguish truth and deception at an 85% level of accuracy (<http://www.reid.com/service-bai-interview.html>).

Unfortunately, psychological research has generally failed to support the claim that individuals can attain high levels of performance in making judgments of truth and deception. Over the years, numerous studies have demonstrated that individuals perform at no better than chance level in detecting deception (DePaulo, Stone, & Lassiter, 1985), that training programs produce only inconsistent improvements in performance compared with control conditions (Bull, 1989; Kassin & Fong, 1999; Porter, Woodworth, & Birt, 2000; Vrij, 1994; Zuckerman, Koestner, & Alton, 1984; Zuckerman, Koestner, & Colella, 1985), and that police investigators and others with relevant on-the-job experience perform only slightly better, if at all (Bull, 1989; DePaulo, 1994; DePaulo & Pfeifer, 1986; Ekman & O'Sullivan, 1991; Ekman, O'Sullivan, & Frank, 1999; Koehnken, 1987; Porter et al., 2000). Thus, although many in the law enforcement community assume, often with great confidence, that investigators can use verbal and nonverbal behavioral cues to make accurate judgments of truth and deception, there is little hard evidence to support this assumption.

Unfortunately, the implications of falsely inferring deceit in the interrogation room can be quite costly to an innocent suspect. Consider, for example, the military trial of *U.S. v. PFC Timothy Bickel* (1999), in which one of us testified as an expert witness (the second author). In this case, a confession to rape was extracted by a combination of five agents who used persistent and highly aggressive techniques (e.g., explicit and implied promises and threats, negative feedback on a polygraph test that was described as infallible, and the use of minimization) despite the absence of independent evidence against the defendant. When asked why they chose to interrogate Bickel and not others, one investigator noted that he showed "signs of deception based on the training we have received." More specifically, this investigator stated, "His body language and the way he reacted to our questions told us that he was not telling the whole truth. Some examples of body language is that he tried to remain calm but you could tell he was nervous and every time we asked him a question his eyes would roam and he would not make direct contact, and at times he would act pretty sporadic and he started to cry at one time." Correctly, we believe, this defendant was acquitted at trial. There are many other instances too, such as the case of Tom Sawyer, in Florida, in which investigators accused the defendant of sexual assault and murder, interrogated him for 16 hr, issued threats, and extracted a confession likely to have been false (cited in Ofshe & Leo, 1997). The confession was suppressed by the trial judge, and, in the absence of other evidence,

the charges were dropped. Most intriguing to us was the manner in which Sawyer became a prime suspect—his face flushed and he appeared embarrassed during an initial interview, a behavioral reaction interpreted as a sign of deception. However, what the investigators did not know was that Sawyer was a recovering alcoholic with a social anxiety disorder that caused him to sweat profusely and blush in evaluative social situations.

Given the profound implications for error at this early stage of criminal investigations, it is important to understand the influence of training and prior experience on the deception-detection performance of police investigators. Regrettably, however, a quick survey of the literature may itself prove somewhat deceptive, as many studies have reported only global accuracy rates (i.e., percentage correct) across all targets. We believe that such a method of analysis may fall short of providing a complete account of the performance of individuals as a function of levels of experience or training. Rather, considering the influence of such manipulations across both *truthful* and *deceitful* targets would allow greater insight into processes underlying deception-detection decisions, particularly when placed within the context of signal detection theory (SDT; Green & Swets, 1966; MacMillan & Creelman, 1991). Unfortunately, in only a few cases have researchers actually provided such estimates of truthful and deceitful performance (Ekman et al., 1999; Kassin & Fong, 1999; Porter et al., 2000; Vrij, 1993; Vrij & Mann, in press), and in no cases have researchers directly applied SDT to assessing deception-detection performance.

Although it is possible that training or experience may increase an individual's ability to discriminate between truthful and deceptive target persons, it may alternatively bias decisions toward truth or deceit on the basis of a variety of factors. As conceived by Green and Swets (1966), SDT provides a framework for separating performance into two conceptually and computationally independent parameters, namely *discrimination accuracy*—the ability of an individual to correctly detect a signal (deception) versus correctly reject its absence (truth), and *response bias*—the degree of evidence necessary for the individual to respond that a signal (deception) has been presented. Although SDT has been widely used throughout psychology, as in the study of decision making among air-traffic controllers, doctors, and clinical psychologists, it has not been applied to the study of deception-detection judgments (but see Thompson, 1978, regarding the detection of social cues). The implication is that various targets, expectations, experiences, and other manipulations thus far presented in the deception literature may differentially influence the two SDT parameters. Knowing how these parameters are influenced should also provide insight into the nature of the psychological effects underlying the various manipulations.

With regard to the effects of training or prior experience, proponents seem to suggest that certain manipulations would only influence discrimination accuracy by enhancing one's ability to correctly differentiate truth from deceit. This certainly represents the aim of most training methods that seek to provide knowledge of verbal and nonverbal cues inherent in deception. On the basis of the relative independence of the two SDT parameters (cf. Snodgrass & Corwin, 1988), it may also be assumed that training and experience would *not* influence the response

bias parameter, that is, the individual's tendency to overly respond "truth" versus "deceit."

### EFFECT SIZE ANALYSIS OF PREVIOUS DECEPTION-DETECTION RESEARCH

We first evaluated this hypothesis by conducting a brief survey and reanalysis of existing research. In particular, we looked for studies that examined the influence of either experience, such as law enforcement, or specific training in the detection of deception. Our literature search was constrained by the requirements that studies (a) report accuracy for *both* truthful and deceitful targets and (b) include a no-training or no-experience control condition. For studies in which only gross accuracy estimates were reported, authors were contacted for the additional relevant information. Overall, we were able to locate four studies that examined the influence of prior experience (Chahal & Cassidy, 1995; DePaulo & Pfeifer, 1986; Ekman et al., 1999; Porter et al., 2000) and two that assessed the influence of training on deception-detection performance (Kassin & Fong, 1999; Koehnken, 1987). Together, these six studies represented the responses of 1,161 participants. Other studies were excluded either because of insufficient reporting of data (deTurck, 1991; deTurck & Miller, 1990; Ekman & O'Sullivan, 1991; Kraut & Poe, 1980; Vrij, 1994; Zukerman et al., 1984, 1985) or because they did not include a control condition (Vrij, 1993; Vrij & Graham, 1997; Vrij & Mann, 2001). A studywise SDT analysis (MacMillan & Creelman, 1990; MacMillan & Kaplan, 1985) was performed on the performance means that were either reported in the paper or provided by the authors. Estimates of *hits* (i.e., the proportion of deceptive persons correctly identified as such) and *false alarms* (i.e., the proportion of truthful persons incorrectly identified as deceptive) were used to compute studywise SDT estimates of both discrimination accuracy ( $A'$ ) and response bias ( $B''_D$ ) for each condition (see Donaldson, 1992). Finally, the difference in performance between the experimental and control conditions was computed by estimating the variance for each SDT parameter and computing Cohen's  $d$  effect size (see Rosenthal, 1994). Positive effect sizes for  $A'$  would indicate that participants in the experimental condition (trained or experienced) outperformed the controls, whereas positive effect sizes for  $B''_D$  would indicate that participants in the experimental condition exhibited a more conservative response bias (i.e., higher likelihood of responding "truthful" across all cases) than did controls.

Studywise estimates of discrimination accuracy ( $A'$ ), response bias ( $B''_D$ ), and effect size (Cohen's  $d$ ) are presented in Table 1. Across studies, the average weighted effect size for the  $A'$  measure was Cohen's  $d = -0.129$ , a non-significant effect size,  $z = 1.92$ , with confidence intervals  $(-0.261, 0.003)$ . In contrast, the  $B''_D$  weighted effect size was significant, Cohen's  $d = -0.317$ ,  $z = 4.75$ ,  $p < .001$ , with confidence intervals  $(-0.448, -0.186)$ . Thus, across studies, both training and prior experience engendered a more liberal response criterion (i.e., a bias toward responding "deceitful") when compared to participants in the no-training and no-experience control conditions. In contrast to what one might have expected, however, training and experience had no significant effect on participants' ability to accurately discriminate truth from deceit.

**Table 1.** Estimates of Signal Detection Performance ( $A'$  and  $B''_D$ ) and Effect Sizes (Cohen's  $d$ ) for the Influence of Experience or Training in a Deception-Detection Task

Study	Manipulation	$N$	$A'$		$B''_D$	
			$M$	Cohen's $d$	$M$	Cohen's $d$
Porter et al. (2000)	Control vs. parole officers	32 vs. 32	0.31	-0.867	0.02	-1.050
Ekman et al. (1999)	Control vs. police officers	425 vs. 105	0.10	-0.005	-0.60	-0.342
			0.70		-0.04	
DePaulo & Pfeifer (1986)	Control vs. police officers	161 vs. 258	0.69	-0.187	-0.24	-0.208
			0.59		0.47	
Chahal & Cassidy (1995)	Control vs. social workers	40 vs. 20	0.54	-0.041	0.35	-0.576
			0.75		0.16	
Kassin & Fong (1999)	Naïve vs. trained	20 vs. 20	0.73	-0.451	-0.18	-0.547
			0.63		0.14	
Kohnken (1987)	Naïve vs. trained	20 vs. 60	0.52	-0.159	-0.18	-0.203
			0.41		-0.60	
			0.37		-0.48	

### EMPIRICAL VALIDATION OF THE "INVESTIGATOR BIAS" EFFECT

Our effect size analysis of the existing literature yielded an *investigator bias effect*, suggesting that training and prior experience lead to a perceptual bias toward judgments of deceit. As this result was unexpected, and as previous studies had neither computed signal detection parameters nor predicted (a priori) such a finding, we attempted to assess directly the validity of this investigator bias in a sample of law enforcement officers. We also sought to extend the ecological validity of the deception-detection task to more closely match the kinds of forensic judgments that investigators routinely make. Of the studies that were included in our studywise analysis, only Kassin and Fong's paradigm utilized stimuli that simulated an investigative interview (Kassin & Fong, 1999). Other studies employed either (a) the classic deceptive opinion/attitude paradigm (DePaulo & Pfeifer, 1986; Ekman et al., 1999), (b) deceptive witnesses to an event (Chahal & Cassidy, 1995; Kohnken, 1987), or (c) deceptive accounts of personal experiences (Porter et al., 2000). Although some previous studies have developed more realistic, forensically relevant detection tasks, these studies either did not include the necessary control condition (e.g., Vrij, 1993) or failed to provide the statistics necessary to examine an investigator bias effect (e.g., Kraut & Poe, 1980). As noted earlier, our concern is that if investigators are truly biased toward deception (or guilt) in their initial judgments, then preinterrogation interviews could prompt the subsequent use of strong, pressure-filled methods of interrogation that, in turn, could increase the risk of coerced false confessions.

To provide a true empirical test of the investigator bias hypothesis, we assessed the deception-detection abilities of police investigators in the context of an investigative interview. In doing so, the present investigation extended recent work by Kassin and Fong (1999), who experimentally trained some student participants, but not others, in the detection of truth and deceit, before obtaining judgments of mock suspects. Their study was unique in two ways. First, the authors administered the popular "Reid Technique," which has been employed in the training of tens of thousands of law enforcement personnel. This training purports to enhance deception-detection ability through the use of various verbal and nonverbal behavioral cues (see Inbau

et al., 2001). Second, these authors created a set of forensically relevant stimulus tapes depicting brief interviews and denials of individuals who were either truly guilty or innocent of committing one of four mock crimes.

Kassin and Fong (1999) showed the videotapes to trained and naïve student participants who attempted to detect whether the mock suspects were lying or truthful. Their results indicated that although training did not increase overall detection accuracy, it did increase the confidence that trained students had in their judgments as well as the number of reasons they cited as a basis for their judgments. As shown in Table 1, our reanalysis of the data also indicated that the training procedure triggered a response bias toward guilt. From a practical standpoint, of course, these data are importantly limited by the fact that the observers were college students, not police detectives, and that the training manipulation was brief and condensed, not offered in the context of professional specialization. As such, this study was designed to extend this paradigm by testing experienced police investigators, some of whom had received training in interviewing, interrogation, and the detection of deception. Our objectives were twofold: (1) to compare the judgments of police investigators to Kassin and Fong's trained and naïve college students (Kassin & Fong, 1999) and (2) to examine—within our samples of investigators—the correlations between experience, training, and various indices of performance.

## METHOD

### Participants

Forty-four North American law enforcement investigators participated in the study. Twenty-five investigators were affiliated with local police departments in Florida, and 19 were from local departments in Ontario, Canada. No differences were found in the experience or performance of the two samples. Investigators averaged 13.7 years of law enforcement experience ( $SD = 6.5$ ), and 68% had undergone formal professional training in interviewing, interrogation, and the detection of the deception.

### Materials

Investigators were shown Kassin and Fong's crime interview tapes, the details of which are provided in their study (Kassin & Fong, 1999, pp. 501–504). Altogether, eight guilty suspects committed one of four mock crimes (vandalism, shoplifting, breaking and entering, a computer break-in), and eight innocent suspects were instructed merely to appear at the scenes of these crimes. All suspects were apprehended, taken to an interrogation room, and questioned by an adult male playing the role of a detective. The interrogator, "Detective McCarthy," was informed about the mock crime that was committed but was blind to each suspect's guilt or innocence. The interviews, which were videotaped, ranged from 3.5 to 6.0 min in duration, with a mean of 4 min 35 s. Self-reports of the participant suspects indicated that the sessions were somewhat stress-provoking, and that the detective was intimidating in his demeanor.

**Table 2.** Deception-Detection Performance of Students and Police Investigators

	Kassin & Fong (1999)		
	Naïve students ( <i>N</i> = 20)	Trained students ( <i>N</i> = 20)	Police investigators ( <i>N</i> = 44)
Judgment accuracy	56% (15%)	46% (17%)	50% (19%)
Confidence	5.91 (0.78)	6.55 (1.02)	7.05 (0.96)
No. of reasons cited	2.98 (0.82)	3.96 (0.85)	2.14 (0.71)
Hits	56% (17%)	45% (17%)	62% (21%)
False alarms	42% (34%)	60% (38%)	67% (26%)
<i>A'</i>	0.63 (0.23)	0.52 (0.24)	0.48 (0.28)
<i>B<sub>D</sub>'</i>	0.14 (0.72)	-0.18 (0.80)	-0.46 (0.55)

*Note.* Standard deviations are provided in parentheses.

### Procedure

In both samples, investigators were randomly assigned to view one of two videotapes, each depicting eight male suspects (four guilty and four innocent) repeatedly denying their involvement in the crime. After completing a brief background questionnaire, investigators were instructed that they would view a tape of eight male suspects, and that they were to decide whether each individual was lying or truthful regarding his involvement. As in Kassin and Fong's study, investigators were told that between one fourth and three fourths of the suspects were lying to protect themselves.<sup>5</sup> In addition to making a truth/deception judgment regarding each suspect, investigators rated their confidence on a scale from 1 (*not at all confident*) to 10 (*extremely confident*), and wrote in their own words the basis for their judgment. Given that Kassin and Fong's student participants completed the same task under identical circumstances, their performance will be compared with that of our police investigators.

### RESULTS

Table 2 presents the results for our police investigators as well as those of Kassin and Fong's naïve and trained student participants who responded to the same stimuli (Kassin & Fong, 1999). A one-way analysis of variance (ANOVA) was performed on each performance variable for the three-way contrast of naïve students versus trained students versus police investigators. Planned comparisons then assessed statistical differences in the performance of police investigators to that of the student samples. The  $p < .05$  convention for significance was employed across all analyses.

<sup>5</sup>This range of deception was provided to capture variability in the response bias parameter, and is a generally accepted detection task instruction. Any systematic fluctuations in response bias across conditions resulting from such an instruction would demonstrate the influence of differential strategies or preconceived base rates. Alternatively, overt specification of the 50% deception used in this study would have suppressed the response bias parameter and hindered any such variation across conditions.

### Overall Judgment Accuracy, Confidence, and Reasons Cited

So that our results would be comparable with previous studies, we computed judgment accuracy scores across all targets. As shown in previous studies, overall accuracy (including both truthful and deceitful targets) for the investigators and student participants did not significantly differ,  $F(2, 81) = 1.54$ ,  $ns$ ,  $\eta^2 = .03$ . However, a significant effect on confidence was observed,  $F(2, 81) = 11.50$ ,  $p < .001$ ,  $\eta^2 = .22$ . Planned comparisons revealed that both naïve and trained students were significantly less confident than police investigators,  $t(62) = 4.87$  and  $2.16$ ,  $ps < .001$  and  $.05$ , respectively. Interestingly, judgment accuracy and confidence were not significantly correlated across investigators,  $r(43) = .20$ ,  $ns$ , indicating poor calibration in detection performance. Participants were also asked to indicate their reasons for judging the target as truthful versus deceitful. Results revealed a significant effect across participant samples,  $F(2, 81) = 39.25$ ,  $p < .001$ ,  $\eta^2 = .49$ . Planned comparisons indicated that both naïve and trained students produced significantly more reasons than did police investigators,  $t(62) = 4.15$  and  $8.95$ , respectively, all  $ps < .001$ . Results of a comparative analysis of the combined sample of students versus police investigators also replicated the above pattern.

### Signal Detection Analysis

As described earlier, we separated performance into estimates of “hits” (the proportion of deceitful suspects correctly identified as deceitful) and “false alarms” (the proportion of truthful suspects incorrectly identified as deceitful). Analysis of hit responses yielded a significant effect across participant samples,  $F(2, 81) = 5.38$ ,  $p < .01$ ,  $\eta^2 = .12$ . Planned comparisons revealed that trained students produced significantly fewer hits relative to the police investigators,  $t(62) = 3.18$ ,  $p < .01$ ; naïve students, however, did not significantly differ from the investigators,  $t(62) = 1.13$ ,  $ns$ . A significant main effect of participant sample was also observed with false alarm responses,  $F(2, 81) = 4.58$ ,  $p < .01$ ,  $\eta^2 = .10$ . Planned comparisons revealed that naïve students generated significantly fewer false alarms relative to police investigators,  $t(62) = 3.25$ ,  $p < .01$ , although the difference between trained students and investigators was not significant,  $t(62) = 0.88$ ,  $ns$ . When student samples were combined, results indicated that investigators generated significantly more hits,  $t(82) = 2.71$ ,  $p < .01$ , and more false alarms,  $t(82) = 2.00$ ,  $p < .05$ .

When these estimates were used to compute measures of  $A'$  and  $B''_D$ , results indicated a significant investigator bias effect, consistent with our studywise analysis. Analysis of the discrimination accuracy parameter revealed no significant differences across participant samples,  $F(2, 81) = 2.11$ ,  $ns$ ,  $\eta^2 = .05$ . Yet a significant effect of response bias was observed,  $F(2, 81) = 5.92$ ,  $p < .01$ ,  $\eta^2 = .13$ . Planned comparisons revealed that naïve students significantly differed from police investigators,  $t(62) = 3.68$ ,  $p < .001$ . Although trained students demonstrated a more conservative response bias relative to police investigators, this difference failed to reach conventional levels of significance,  $t(62) = 1.68$ ,  $ns$ . A comparison of the combined sample of students versus investigators also indicated no significant differences on discrimination accuracy,  $t(82) = 1.61$ ,  $ns$ , but an investigator bias effect in estimates of response criterion,  $t(82) = 3.05$ ,  $p < .01$ .



### Correlational Analyses

Comparison of law enforcement professionals to college students is one way to gauge the role of experience and training on deception-detection performance. Within our sample of investigators, we also examined the correlations between the amount of training and experience the investigators had received and key measures of task performance. First, prior experience (years) in law enforcement was not related to either overall judgment accuracy,  $r(44) = -.19$ , decision confidence,  $r(44) = -.07$ , or the number of reasons investigators cited for their decisions,  $r(44) = -.03$ . In terms of signal detection parameters of performance, however, greater prior experience was significantly correlated with both an increase in false alarm responses,  $r(44) = .33$ ,  $p < .05$ , and an increased bias in responding "deceit,"  $r(44) = -.34$ ,  $p < .05$ . A similar pattern emerged for the effect of deception-detection training with regard to its lack of influence on both judgment accuracy and decision confidence,  $r(44) = -.04$  and  $-.01$ , respectively. As might have been expected, however, prior training did significantly increase the number of reasons that investigators cited for their decisions,  $r(44) = .41$ ,  $p < .01$ , and marginally prompted an increase in false alarm responses,  $r(44) = .25$ ,  $p < .10$ , and a bias to see deceit,  $r(44) = -.28$ ,  $p < .10$ .

### DISCUSSION

While detecting deception seems an inherently difficult task, the judgments that police investigators make of suspects on the basis of preinterrogation interviews represent a critical choice point in law enforcement, as they determine whether suspects are sent home or subjected to guilt-presumptive interrogation. Given the importance of this decision-making process, deception-detection techniques that rely on an analysis of verbal and nonverbal behavioral cues have been used as a basis for training. Despite the assumption that such programs enhance accuracy, however, published psychological research has generally failed to demonstrate performance increments as a function of special training or prior law enforcement experience (Bull, 1989; DePaulo & Pfeifer, 1986).

Our goal was to better understand the influence of training and experience on deception-detection judgments. Unfortunately, previous studies had focused upon global accuracy rates rather than separating performance for both truthful and deceitful targets. By separating these two aspects of performance, the application of SDT affords the opportunity to distinguish between two conceptually and computationally independent parameters of detection skill: *discrimination accuracy*—correctly detecting a signal (deception) versus correctly rejecting its absence (truth), and *response bias*—the threshold of evidence necessary for an individual to respond that a signal (deception) has been presented. On the basis of the assumptions underlying manipulations of training or experience, one would expect that discrimination accuracy would be enhanced; however, given the relative independence of the two parameters, response bias should not be influenced by such manipulations.

To test this hypothesis, we first conducted a studywise analysis of the previous literature in which estimates of discrimination accuracy and response bias were calculated. Effect sizes were computed for each study, after which they were weighted

by sample size and aggregated across studies for each SDT parameter. The results of this analysis indicated a reliable effect on estimates of response bias. In particular, training and experience appeared to loosen participants' response criterion, thereby increasing the likelihood that they would judge targets as deceitful rather than truthful relative to a no-training or experience control group. Across studies, however, training and experience produced no reliable effects on the ability to discriminate between truth and deceit.

To more directly assess the validity of this "investigator bias effect" in a forensically relevant setting, we presented a sample of law enforcement investigators with videotapes of brief interviews in which guilty or innocent male suspects denied their involvement in a mock crime. The task was to determine whether each suspect was truthful in his denial or lying to cover his guilty actions. We compared the performance of our police investigators with that of naïve and trained student participants from the study of Kassin and Fong (1999), who completed the task with the same stimuli and under identical conditions. Overall, the investigators did not outperform Kassin and Fong's trained or naïve student participants (Kassin & Fong, 1999), in terms of either global accuracy or discrimination performance (*A'*). However, they were significantly more confident in their judgments and significantly more likely to respond "deceitful" rather than "truthful"—a liberal response bias that is consistent with our reanalysis of the literature. In short, the pivotal decision investigators must make regarding whether to further interrogate a suspect may be based on prejudgments of guilt, confidently made, but frequently in error.

One might argue that our investigators were limited in their performance by the relative brevity and low-stakes nature of the taped interviews. As noted earlier, however, their accuracy rate is consistent with past studies of law enforcement professionals. Moreover, this possible limitation with regard to the targets of judgment used in this study cannot explain the response bias or high confidence levels exhibited by investigators relative to students (indeed, from a metacognitive perspective, one would hope that professionals would adjust their confidence levels according to such perceived limitations). One might also argue that investigators were limited by their ability to merely observe, not actively conduct, the interviews. Importantly, however, research in other contexts has shown that judgments of truth and deception are more accurate when made by observers than by conversational interactants (Buller, Strzyzewski, & Hunsaker, 1991).

Given the potential and serious consequences of an investigator bias effect, it will be important in future studies to consider both the theoretical mechanisms and practical implications of the observed bias. From a theoretical perspective, social cognitive mechanisms may be responsible for the investigator bias due to experience or training. For example, investigators in a deception-detection task may be influenced by subjective probabilities, or base rates, for the incidence of truth versus deceit (Fiedler, 2000), and these probabilities may be influenced by factors, such as the race, age, ethnicity, criminality, or status of the suspect (Ruby & Brigham, 1996), or by the contextual basis of the interview itself (e.g., whether there is independent, corroborative evidence). These initial attributions may then set the stage for other mechanisms, including the manner in which investigators encode and process

seemingly relevant information (Kassin et al., 2001; Nickerson, 1998). For example, police investigators may encode and weight information that has little discriminative validity (see Simon, 1991). Although such mechanisms were not tested in this study, future research should consider their influence in the investigator bias effect and seek to dissociate the effects of training and prior experience.

From a practical standpoint, the investigator bias uncovered in this study suggests that many innocent suspects are in a precarious situation before undergoing formal interrogation. Routinely, investigators seek to assess a suspect's veracity in order to determine whether to launch into an interrogation, a process that, at times, elicits coerced false confessions from innocent people. Another possible detrimental effect on the innocent suspect may arise from the influence of trial testimony provided by investigative officers on the perceived truthfulness of the defendant. Although this proposition remains to be empirically tested, it is possible that such confident and descriptive testimony—accurate or not, and biased or not—may have a strong effect on jurors and jury decision-making. Thus, in light of the profound risks to the accused, both in the interrogation room and subsequently in the courtroom, we believe that further research is warranted regarding the theoretical mechanisms responsible for the effect and the practical ramifications on the disposition of a given case.

### ACKNOWLEDGMENTS

The authors thank John Turtle for assisting with data collection.

### REFERENCES

- Bull, R. (1989). Can training enhance the detection of deception? In J. C. Yuille (Ed.), *Credibility assessment* (pp. 83–99). London: Kluwer Academic Publishers.
- Buller, D. B., Strzyzewski, K. D., & Hunsaker, F. G. (1991). Interpersonal deception: II. The inferiority of conversational participants as deception detectors. *Communication Monographs, 58*, 25–40.
- Chahal, K., & Cassidy, T. (1995). Deception and its detection in children: A study of adult accuracy. *Psychology, Crime and Law, 1*, 237–245.
- 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. I., & Lassiter, G. 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: Effects of gender. *Communication Reports, 4*, 81–89.
- deTurck, M. A., & Miller, G. R. (1990). Training observers to detect deception: Effects of self-monitoring and rehearsal. *Human Communication Research, 16*, 603–620.
- Donaldson, W. (1992). Measuring recognition memory. *Journal of Experimental Psychology: General, 121*, 275–277.
- 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.
- Fiedler, K. (2000). Beware of samples! A cognitive–ecological sampling approach to judgment biases. *Psychological Review, 107*(4), 659–676.
- Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

- Gudjonsson, G. H. (1992). Interrogative suggestibility: Factor analysis of the Gudjonsson Suggestibility Scale (GSS 2). *Personality and Individual Differences, 13*(4), 479–481.
- Inbau, F. E., Reid, J. E., Buckley, J. P., & Jayne, B. C. (2001). *Criminal interrogation and confessions* (4th ed.). Gaithersburg, MD: Aspen Publishers.
- Kassin, S. M. (1997). The psychology of confession evidence. *American Psychologist, 52*(3), 221–233.
- Kassin, S. M., & Fong, C. T. (1999). "I'm innocent!": Effects of training on judgments of truth and deception in the interrogation room. *Law and Human Behavior, 23*, 499–516.
- Kassin, S. M., Goldstein, C. J., & Savitsky, K. (2001). *Behavioral confirmation in the interrogation room: On the dangers of presuming guilt*. Unpublished manuscript.
- Koehnken, G. (1987). Training police officers to detect deceptive eyewitness statements: Does it work? *Social Behavior, 2*, 1–17.
- 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.
- Leo, R. A., & Ofshe, R. J. (1998). The consequences of false confessions: Deprivations of liberty and miscarriages of justice in the age of psychological interrogation. *Journal of Criminal Law and Criminology, 88*(2), 429–496.
- Macmillan, N. A., & Creelman, C. D. (1990). Response bias: Characteristics of detection theory, threshold theory, and "nonparametric" indexes. *Psychological Bulletin, 107*, 401–413.
- MacMillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide*. New York: Cambridge University Press.
- MacMillan, N. A., & Kaplan, H. L. (1985). Detection theory analysis of group data: Estimating sensitivity from average hit and false alarm rates. *Psychological Bulletin, 98*, 185–199.
- Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology, 2*, 175–220.
- Ofshe, R. J., & Leo, R. A. (1997). The social psychology of police interrogation: The theory and classification of true and false confessions. *Studies in Law, Politics, and Society, 16*, 189–251.
- Porter, S., Woodworth, M., & Birt, A. R. (2000). Truth, lies, and videotape: An investigation of the ability of federal parole officers to detect deception. *Law and Human Behavior, 24*, 643–658.
- Radelet, M. L., Bedau, H. A., & Putnam, C. E. (1992). *In spite of innocence: Erroneous convictions in capital cases*. Boston: Northeastern University Press.
- Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 111–124). New York: Russell Sage Foundation.
- Ruby, C. L., & Brigham, J. C. (1996). A criminal schema: The role of chronicity, race, and socioeconomic status in law enforcement officials' perceptions of others. *Journal of Applied Social Psychology, 26*, 95–112.
- Scheck, B. C., Neufeld, P., & Dwyer, J. (2000). *Actual innocence: Five days to execution and other dispatches from the wrongly convicted*. New York: Doubleday.
- Simon, D. (1991). *Homicide: A year on the killing streets*. New York: Ivy Books.
- Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General, 117*, 34–50.
- Thompson, S. C. (1978). Detection of social cues: A signal detection theory analysis. *Personality and Social Psychology Bulletin, 4*, 452–455.
- Vessel, D. (1998, October). Conducting successful interrogations. *FBI Law Enforcement Bulletin, 1*–6.
- Vrij, A. (1993). Credibility judgments of detectives: The impact of nonverbal behavior, social skills, and physical characteristics on impression formation. *Journal of Social Psychology, 133*, 601–610.
- Vrij, A. (1994). The impact of information and setting on detection of deception by police detectives. *Journal of Nonverbal Behavior, 18*, 117–132.
- Vrij, A., & Graham, S. (1997). Individual differences between liars and the ability to detect lies. *Expert Evidence, 5*, 144–148.
- Vrij, A., & Mann, S. (2001). Who killed my relative? Police officers' ability to detect real-life high-stake lies. *Psychology, Crime and Law, 7*, 119–132.
- Vrij, A., & Mann, S. (2001). Telling and detecting lies in a high-stake situation: The case of a convicted murderer. *Applied Cognitive Psychology, 15*, 187–203.
- Zuckerman, M., Koestner, R., & Alton, A. O. (1984). Learning to detect deception. *Journal of Personality and Social Psychology, 46*, 519–528.
- Zuckerman, M., Koestner, R., & Colella, M. J. (1985). Learning to detect deception from three communication channels. *Journal of Nonverbal Behavior, 9*, 188–194.