Law and Human Behavior, Vol. 29, No. 2, April 2005 (© 2005)
DOI: 10.1007/s10979-005-2416-9

"I'd Know a False Confession if I Saw One": A Comparative Study of College Students and Police Investigators

Saul M. Kassin,[1,4] Christian A. Meissner,[2] and Rebecca J. Norwick[3]

College students and police investigators watched or listened to 10 prison inmates confessing to crimes. Half the confessions were true accounts; half were false—concocted for the study. Consistent with much recent research, students were generally more accurate than police, and accuracy rates were higher among those presented with audiotaped than videotaped confessions. In addition, investigators were significantly more confident in their judgments and also prone to judge confessors guilty. To determine if police accuracy would increase if this guilty response bias were neutralized, participants in a second experiment were specifically informed that half the confessions were true and half were false. This manipulation eliminated the investigator response bias, but it did not increase accuracy or lower confidence. These findings are discussed for what they imply about the post-interrogation risks to innocent suspects who confess.

KEY WORDS: confessions; deception; police.

In recent years, numerous high-profile DNA exonerations have surfaced, leading social science researchers, legal scholars, policy makers, and the news media to revisit the evidence upon which innocent people had been prosecuted, convicted, and imprisoned. As reported in Scheck, Neufeld, and Dwyer's (2000) *Actual Innocence,* and as confirmed by data that have accumulated since that time, 20–25% of DNA exoneration cases contained full or partial confessions in evidence (www.innocenceproject.org). The shocking exonerations in New York's Central Park jogger case illustrate the point. In 1989, a female jogger was raped, brutally beaten, and left for dead in Central Park. Within 72 h, five juveniles, 14–16 years old, confessed to the assault in lurid detail. Four of the confessions were videotaped. The boys immediately retracted their statements, claiming that they were coerced

[1]Williams College.
[2]Florida International University.
[3]Harvard University.
[4]To whom correspondence should be addressed at Department of Psychology, Williams College, Williamstown, Massachusetts 01267; e-mail: skassin@williams.edu.

0147-7307/05/0400-0211/1 © 2005 Springer Science+Business Media, Inc.

and false. Yet solely on the basis of these statements, they were convicted by juries and sentenced to prison. Thirteen years later, Matias Reyes—an imprisoned serial rapist and murderer—confessed that he alone had attacked the jogger. The Reyes confession, unlike those of the boys, was corroborated by DNA tests of semen found at the crime scene. Apparently, despite the spotlight cast by the national news media, this one case contained five false confessions (Kassin, 2002; Saulny, 2002; Morgenthau, 2002).

The jogger case and others involving proven false confessions point to two problems. The first is that innocent people can be induced to confess to crimes they did not commit. Over the years, psychologists have proposed theories of motivation, decision-making, and social influence to understand the processes of interrogation, and have used an array of research methods to understand how and why certain interrogation tactics lead suspects to confess (Davis & O'Donohue, 2003; Drizin & Leo, 2004; Gudjonsson, 1992, 2003; Hilgendorf & Irving, 1981; Kassin, 1997; Kassin & Wrightsman, 1985; Lassiter, 2004; Leo, 1996; Leo & Ofshe, 1998; Redlich & Goodman, 2003; Wrightsman & Kassin, 1993; Zimbardo, 1967). There is, however, a second problem evident in the jogger case and others like it: that police, district attorneys, judges, and juries believed these confessions, indicating perhaps that they cannot distinguish between self-incriminating statements that are true and those that are false. One could argue that interrogation is psychologically coercive, and that innocent people sometimes confess, but that such errors will ultimately be detected by authorities and corrected. Essential to this presumed safety net is the commonsense assumption that "I'd know a false confession if I saw one."

Is there a reason to believe that investigators can accurately distinguish between true and false confessions? Consistently, research has shown that people are not proficient at judging truth and deception, often performing at no better than chance levels (DePaulo, Lassiter, & Stone, 1982; Memon, Vrij, & Bull, 2003; Vrij, 2000), that training programs produce only small and unreliable improvements in performance (Bull, 1989; Kassin & Fong, 1999; Porter, Woodworth, & Birt, 2000; Vrij, 1994; Zuckerman, Koestner, & Alton, 1984), and that police and other detection deception "professionals" typically perform no better than laypeople when such comparisons are made (Bull, 1989; DePaulo, 1994; DePaulo & Pfeifer, 1986; Ekman & O'Sullivan, 1991; Ekman, O'Sullivan, & Frank, 1999; Garrido & Masip, 1999; Garrido, Masip, & Herrero, 2004; Koehnken, 1987; Porter et al., 2000). In short, the law enforcement community assumes that investigators can become highly accurate judges of truth and deception (Inbau, Reid, Buckley, & Jayne, 2001), but there is little if any evidence to support this claim (for a recent meta-analysis of presumed cues to deception, see DePaulo et al., 2003; for a comprehensive review of deception detection issues in a forensic context, see Granhag & Strömwall, 2004).

To address this question in a criminal context, Kassin and Fong (1999) examined whether people can distinguish true and false *denials*—and whether police training in the use of verbal and nonverbal deception cues would increase the accuracy of such judgments. In Phase 1, participants committed one of four mock crimes and then denied their involvement in an interview. In Phase 2, observers were either trained in the Reid technique approach to deception detection or not trained

before judging these taped interviews. As in other research, the results of this study indicated that observers could not significantly distinguish between the truthful and deceptive suspects. In fact, those who underwent the training were both less accurate and more confident than naïve controls. In a follow-up study, Meissner and Kassin (2002) showed these interviews to experienced detectives and found that although they were not more accurate than students, they were more confident—and more likely to make false positive errors, illustrating an "investigator bias" toward perceiving deception. The pivotal decision of whether or not to interrogate a suspect is thus based on prejudgments of guilt that are confidently made but biased toward guilt and often in error.

Past research has examined the impact of confession evidence on jurors and others in the criminal justice system. Mock jury studies have shown that people do not adequately discount confession evidence even when it is logically appropriate to do so (Kassin & Wrightsman, 1980, 1985). Indeed, confessions have more impact than eyewitness and character testimony, other powerful forms of evidence (Kassin & Neumann, 1997), and they increase the conviction rate even among mock jurors who see the statements as coerced and who self-report being uninfluenced by them (Kassin & Sukel, 1997). More generally, confessions tend to overwhelm other information, including evidence of innocence, resulting in a chain of adverse legal consequences—from arrest through prosecution, conviction, and incarceration (Drizin & Leo, 2004; Leo & Ofshe, 1998). Thus, to safeguard against wrongful convictions, it is important that confessions be accurately assessed prior to the onset of court proceedings. But can people in general, and law enforcement investigators in particular, distinguish true from false confessions?

This research tested a common assumption, not previously tested, that "I'd know a false confession if I saw one." To examine this question, we conducted a two-phased experiment. First, we recruited male prison inmates to take part in a pair of videotaped interviews—one in which they gave a full confession to the crime for which they were incarcerated, the other in which they concocted a false confession to a crime described by the experimenter that they did not commit. Second, we showed civilian and police observers a stimulus tape of 10 different inmates, each giving a true or false confession to one of five crimes. After each statement, participants judged whether the individual was guilty or innocent and rated their confidence in that judgment.

In addition to developing this novel paradigm for assessing judgments of confessions, this research was designed with three goals in mind. The first was to compare untrained lay observers and police investigators for their judgment accuracy and confidence—and to assess, within the law enforcement sample, the correlations among training in deception detection, years of experience, and performance. Second, we sought to elucidate the nature of the investigator response bias previously found. Research shows that police investigators are generally prone to see deception, which typically signals guilt (Masip, Alonso, Garrido, & Anton, in press; Meissner & Kassin, 2002). But what does this tendency imply when it comes to assessing confessions? In forensic settings, lying and guilt are naturally conflated, as innocent suspects state truthful alibis whereas criminals lie in their denials. By having observers assess true and false confessions rather than denials, so that "truth"

judgments indicate guilt and "false" judgments indicate innocence, we sought to determine whether the disposition among police is to see deception (i.e., by disbelieving confessions) or guilt (i.e., by believing confessions). Third, we sought to examine whether discrimination accuracy in judging confessions is influenced by the medium of their presentation. Many law enforcement professionals are trained to assess suspects by attending to behavioral symptoms, many of which are visual in nature (Inbau et al., 2001). Yet studies have suggested that auditory cues are more diagnostic of truth and deception (e.g., Anderson, DePaulo, Ansfield, Tickle, & Green, 1999; DePaulo et al., 1982). In light of recent policy discussions concerning the electronic recording of interrogations, it was important to compare the performance of lay participants and police investigators who viewed the confessions on videotape to those who merely listened on audiotape.

EXPERIMENT 1

Method

This experiment was conducted in two phases. First, a group of prison inmates provided true and false confessions that were recorded on audiotape and on videotape. Next, confessions were presented for judgment to college students and police investigators.

Participants

Male inmates from a Massachusetts state correctional facility were recruited and paid to take part in a pair of videotaped interviews. The facility houses roughly a thousand state and county offenders. In response to a call for research subjects, a total of 20 inmates volunteered and were paid $20 for their participation. However, one refused to discuss his crime, a second claimed he was innocent, and a third refused to generate a false confession, so statements were obtained from 17 inmates.

One hundred eighteen participants, from two samples, served as judges in Phase 2 of this study. Serving as a convenient sample of laypeople, one group consisted of 61 male and female introductory psychology students who took part in exchange for extra course credit. The second sample consisted of 57 federal, state, and local investigators from Florida and Texas recruited through personal contacts and direct solicitation to their departments. As a group, 47 investigators were male, 10 were female. They had an average of 10.94 years of law enforcement experience, and 58% had received special training in deception detection, interviewing, and interrogation. Within both samples, participants in small group settings were randomly assigned to the videotape or audiotape condition.

The Stimulus Tape

With assistance from prison staff, 20 inmates were recruited and escorted to a special room to take part in the study. Upon arrival, each inmate was seated at a table and introduced to a male interviewer and the female technical assistant who operated the audiovisual equipment. After explaining the task, the interviewer

presented the participant with a written consent form for a signature and read it aloud. This form stated that participants are anonymous ("that my name will not be associated with the results in any way"), that the information they provide is confidential ("to be shared only with others involved in the research project"), that they will be paid $20, and that they may withdraw their consent and discontinue at any time.

Inmates who signed the consent form were next asked to provide a full confession to the crime for which they were in prison, statements that were verified by their records, but not to talk about their arrest, conviction, or incarceration, or other aspects of their recent lives. Specifically, they were instructed: "Tell me about what you did, the crime you committed, that brought you here. Try to give me as much detail as you can about what happened, when, where, who you were with, and so on." To ensure that all stimulus confessions contained the same basic ingredients, each free narrative was followed by a standardized set of 10 questions that probed for who, what, when, where, how, why, and other details, such as: "Had you planned to do it?" "Did anyone see you?" "Afterward, what did you do and where did you go?" "Did you tell anyone about it?" "What did you do with the. . . ?" All sessions were videotaped from a camcorder that was mounted on a tripod behind the interviewer, five feet in front of the inmate. The sessions were also recorded by an audiotape recorder placed on the table.

For a second videotaped interview, each inmate was instructed that, "I'm going to tell you about a crime that you were *not* involved in. I'd like you to lie about it and make up a confession as if you did it. Try to imagine the crime and imagine yourself doing it. Then make up a story filled with details of what happened, what you did, when, where, who you were with, and so on." Each inmate was then given a skeletal, one- or two-sentence description of the true crime described by the preceding participant and offered a couple of minutes to concoct a false confession. As with the true statements, each free narrative was followed by standardized interview questions. Using this yoked design, the first inmate's true confession became the basis of the second inmate's false confession; the second's true confession became the basis of the third's false confession, and so on. The order in which the participants gave true and false confessions was counterbalanced across sessions.

Seventeen inmates provided true and false confessions. However, a number of statements had to be discarded because the inmate, despite instruction, had talked about his arrest, conviction, and incarceration, or strayed out of character (e.g., asking during the statement, "is it okay if I give a made-up name?"). In these instances, the yoked companion confessions had to be discarded as well. Through this procedure, and the elimination of "second" appearances by the same inmate, we created a stimulus videotape and a corresponding audiotape that depicted 10 different individuals, once each, confessing to one of five crimes: aggravated assault, armed robbery, burglary, breaking and entering, and automobile theft and reckless driving. As there is no forensic relevance to the question of whether people can choose between competing confessions, the statements were not explicitly paired for presentation, but the tapes as a whole contained five true confessions and their yoked, false confession counterparts. Except for the constraint that the true and false versions of the same crime not appear in sequence, the 10 confessions were randomized

and presented in a constant order. The entire stimulus tape is 45 min in duration (confessions averaged 4 min, 40 s).

Procedure

Both student and police observers were scheduled and run in small group sessions, and groups were randomly assigned to participate in the videotape or audiotape condition. Before exposure to the taped confessions, all participants were instructed that they would be presented with a number of statements, some that were true others that were false. They were asked not to react publicly or comment on the statements in order to ensure the independence of all responses. They were then handed a 10-page questionnaire, with the pages labeled "Statement 1" through "Statement 10." On each page, one per confession, participants were asked to circle their judgment: "In your opinion, is this individual *guilty* of the crime to which he has confessed, or is he *innocent* of it and telling a false story?" They then rated their confidence in that judgment on a 1–10 point Likert-type scale (1: "not at all confident," 10: "very confident"). At the conclusion of each session, the groups were debriefed and thanked for their participation.

Results

In global judgment accuracy, the results of this study paralleled those obtained for judgments of true and false denials (Meissner & Kassin, 2002). Across participants, conditions, and items, the overall accuracy rate was 53.9%—a level of performance that is both unimpressive and nonsignificant relative to chance performance (z-test for proportions = 0.87). In signal detection terms, the hit rate (the percentage of inmates whose true confessions were correctly identified as true) was 63.6% and the false alarm rate (the percentage of inmates whose false confessions were incorrectly identified as true) was 56.1%. On a 1–10 point scale, the overall mean confidence level was 6.76. Interestingly, judgment accuracy and confidence were negatively correlated (point biserial $r = -.23, p < .02$).

All performance measures were analyzed within a 2 (students, investigators) × 2 (videotape, audiotape) between-subject analysis of variance (ANOVA). On the all-important measure of global accuracy, significant main effects were found for both participant sample and for medium of presentation. Specifically, students were more accurate than investigators (Ms = 58.8 and 48.3%, respectively), $F(1, 114) = 15.49, p < .001, \eta^2 = .12$; and accuracy was greater in the audio than video condition (Ms = 59.3 and 47.8%, respectively), $F(1, 114) = 18.71, p < .001, \eta^2 = .14$. Among the four groups, students in the audiotape condition were the most accurate, exceeding chance level performance ($M = 64.1\%$, z-test for proportions = 1.65, $p < .05$); police investigators in the videotape condition were the least accurate ($M = 42.1\%$, z-test for proportions = .86). The full results within each cell are presented in Table 1.

Students may have been more accurate in their judgments, but police investigators were significantly more confident (Ms = 7.35 and 6.21, respectively), $F(1, 114) = 39.28, p < .001, \eta^2 = .26$. Overall levels of confidence were not affected by medium of presentation (Ms = 6.66 and 6.91 in the audio and video conditions,

Table 1. Key Performance Measures Among Students and Investigators in the Videotape and Audiotape Conditions of Experiment 1

	Students		Investigators	
	Video	Audio	Video	Audio
N	29	32	28	29
Judgment accuracy (%)	53.4	64.1	42.1	54.5
Hit rates (%)	55.9	70.0	57.9	69.7
False alarms (%)	50.3	41.9	73.6	60.7
Confidence	6.18	6.25	7.65	7.06
A'	0.57	0.68	0.39	0.58
B''_{D}	−0.10	−0.21	−0.56	−0.52

respectively), $F(1, 114) = 2.0, p < .20$. Although the difference between participant samples was somewhat larger in the video condition (*M*s = 7.65 vs. 6.18 for investigators and students, respectively) than in the audio condition (*M*s = 7.06 vs. 6.25 for investigators and students), the two-way interaction term was not quite significant, $F(1, 114) = 3.38, p < .07, \eta^2 = .03$.

Using a signal detection framework, we separated performance into estimates of "hits" (the proportion of inmates whose true confessions were correctly identified as true) and "false alarms" (the proportion of inmates whose false confessions were incorrectly identified as true). Analysis of these measures showed that although participant samples did not differ in their hit rates (*M*s = 63.8 and 62.9% for police and students, respectively), $F(1, 114) < 1$, investigators generated significantly more false alarms ($M = 67.1$ and 46.1%, respectively), $F(1, 114) = 28.72$, $p < .001$, $\eta^2 = .20$. Significant main effects were also obtained for medium of presentation. The hit rate was higher in the audio condition ($M = 69.8$ vs. 56.9% in the video condition), $F(1, 114) = 11.50, p < .001, \eta^2 = .09$, and the false alarm rate was higher in the video condition ($M = 62.0$ vs. 51.3% in the audio condition), $F(1, 114) = 7.41, p < .01, \eta^2 = .06$. There were no significant interactions between sample and medium of presentation on these measures (*F*s < 1). When estimates were combined into aggregate measures of discrimination accuracy (A') and response bias (B''_{D}), the results replicated the significant investigator bias effect previously described. Specifically, students exhibited significantly greater discrimination accuracy ($M = .62$ vs. .48), $F(1, 114) = 10.40, p < .002, \eta^2 = .08$, while investigators exhibited a greater response bias toward viewing confessions as true ($M = -.54$ vs. $-.16$), $F(1, 114) = 17.01, p < .001, \eta^2 = .13$. With regard to the medium of presentation, participants exhibited greater discrimination accuracy in the audiotape condition than in the videotape condition, (*M*s = .63 vs. .48), $F(1, 114) = 12.09, p < .001$, $\eta^2 = .10$, but they showed no differences in response bias (*M*s = −.37 vs. −.33), $F(1, 114) = .12$, *ns*.

Comparing students and police investigators is one way to estimate the role of law enforcement training and experience. Another approach is to compare trained and untrained investigators. Within our police sample, we examined the correlations between prior training and experience and key measures of task performance. Overall, 33 out of 57 investigators said they had received special training in deception detection, interviewing, and interrogation. Interestingly, deception detection

training did not significantly correlate with overall accuracy, confidence, or hit rates (rs $= -.13$, .19, and .06, respectively), but it did correlate with the tendency to commit false alarms ($r = .27$, $p < .05$). Hence, while those who were trained did not show less discrimination accuracy ($r = -.15$, $p < .30$), they did exhibit a response bias toward judging confessions as true ($r = -.30$, $p < .05$). With regard to experience, our investigators reported an average of 10.94 years in law enforcement. Measured in this way, experience did not correlate with confidence levels or hit rates (rs $= .04$ and $-.05$, respectively, *ns*), but it did significantly correlate with both overall accuracy ($r = -.26$, $p < .05$) and false alarms ($r = .37$, $p < .005$). Hence, those with more rather than less experience exhibited lower discrimination accuracy ($r = -.35$, $p < .01$) and a greater guilty response bias ($r = -.29$, $p < .05$).

Discussion

In deciding whether to interrogate a suspect, police detectives conduct pre-interrogation interviews in which they make preliminary judgments of truth and deception. Meissner and Kassin (2002) found that while investigators have confidence in their ability to make these judgments, they are no more accurate than laypeople. Moreover, they exhibit a signal detection response bias, tending to judge suspect denials as deceptive. By eliciting judgments of true and false confessions, this study extended previous results in important ways. Once again, investigators were not more accurate than students, only more confident and more biased. Importantly, the response bias currently exhibited reveals that investigators are not disposed to seeing *deception* per se (which, in this study, would mean disbelieving the confessions) but, rather, they are biased toward inferring *guilt* (an inference that involves accepting the confessions as true).

This overall pattern of results concerning judgment accuracy, confidence, and bias has serious implications for the interrogation of innocent suspects and subsequent assessment of their confessions. There are two possible explanations for why police did not distinguish true and false confessions in this study and why they were generally less accurate than naïve college students. One possibility is that law enforcement training and experience introduce systematic bias that reduces overall judgment accuracy (see Meissner & Kassin, 2004). This interpretation is consistent with our internal analyses. It is also not terribly surprising in light of the kinds of deception cues that form the basis for law enforcement training. For example, Inbau et al. (2001) advocate the use of many visual cues—such as gaze aversion, nonfrontal posture, slouching, and grooming gestures—that are not empirically diagnostic of truth or deception (DePaulo et al., 2003). Furthermore, past research has shown that people are more accurate at deception detection when they rely more on such auditory cues as response latency, speech rate, and voice pitch (Anderson et al., 1999; DePaulo et al., 1982). Our results clearly replicated this pattern, with discrimination accuracy significantly higher in the audio than video condition without a significant influence on response bias. In short, it is conceivable that police training in the use of visual cues would impair performance, not improve it.

A second possibility is that investigators' judgment accuracy was compromised by our use of a paradigm in which half of the stimulus confessions were false, a

percentage that is likely far higher than the real world base rate for false confessions. To the extent that law enforcement training and experience leads investigators to presume guilt, and to presume most confessions true, then the response bias they imported from the police station to the laboratory may have proved misleading for a study in which they were told merely that some statements were true and others false. So instructed, investigators judged 65% of the statements to be true, compared to only 55% among student participants, a difference that was highly significant, $t(116) = 3.89$, $p < .001$. Hence, it is possible that investigators performed poorly because of a gross mismatch between the expected and presented base rates for false confessions.

EXPERIMENT 2

To test the hypothesis that judgment accuracy was depressed among investigators relative to students because of differences in base rate expectations, we conducted a second study specifically designed to neutralize the response bias. In this experiment, all participants were shown the 10 videotaped confessions used in Experiment 1, but they were instructed this time that half of the statements were true and half were false. We predicted that this manipulation would neutralize the dispositional response bias of investigators relative to students—and perhaps increase judgment accuracy in the process.

Method

Forty-one participants recruited from two samples judged the confession videotapes in this study. Twenty-one were introductory psychology students (9 male, 12 female), and 20 were state and local police investigators from the state of Florida (15 male, 5 female). Recruited through personal contacts and direct solicitation, the investigators as a group had an average of 11.25 years of law enforcement experience, and 9 (45%) had received special training in deception detection, interviewing and interrogation.

As in the first experiment, all participants took part in small group sessions and judged the same 10 confessions. Prior to watching the tapes, they were admonished not to react overtly and provided with a 10-page questionnaire, with the pages labeled "Statement 1" through "Statement 10." As before, participants were asked after each statement to determine whether the individual was guilty or innocent of the crime for which he had confessed and to rate their confidence in that judgment on a 1–10 point scale. In this experiment, however, they were explicitly told within the instruction that "You will see ten statements. Half are true and half are false."

Results

We sought to eliminate the response bias characteristic of investigators in order to reassess their performance relative to students. In Experiment 1, investigators judged 65.4% of the confessions to be true, compared to 54.6% among students, a difference that was significant, $t(116) = 3.89$, $p < .001$. In this experiment,

however, investigators judged only 51.5% of the statements to be true, compared to 49.5% among students—a difference that was not significant, $t(39) = 1.68$, $p < .11$. The manipulation designed to neutralize the investigator response bias was thus successful.

Across participant samples and items, the overall accuracy rate was 51.2%, a level of performance that did not exceed chance level expectations (z-test for proportions = .26). In signal detection terms, the hit rate was 51.7% and the false alarm rate was 49.3%. On a 1–10 point scale, the overall mean level of confidence was 6.37. As in the first study, there was only a modest, and negative, correlation between judgment accuracy and confidence ($r = -.27$, $p < .10$).

On the measure of global accuracy, students slightly outperformed investigators, but in this study the difference was not significant (Ms = 53.8 and 48.5%, respectively), $t(39) = 1.01$, $p < .50$, and neither group exceeded chance level performance (z-test for proportions = .37 and .09 for students and investigators, respectively, *ns*). Similarly, students and investigators did not differ in their rate of hits (Ms = 53.3 and 50.0%, respectively), $t(39) = .61$, $p > .50$, or false alarms (Ms = 45.7 and 53.0%, respectively), $t(39) = -1.36$, $p < .20$. On the key signal detection measures, the students and investigators did not differ in discrimination accuracy, or A' (Ms = .54 and .46, respectively), $t(39) = 1.06$, $p < .30$, and the previously pronounced response bias (B''_{D}) was no longer significant (Ms = .02 and −.07, respectively), $t(39) = 1.78$, $p < .10$. Yet despite the low and equivalent accuracy rates, and consistent with Experiment 1, investigators were significantly more confident than students in their judgments (Ms = 7.03 and 5.74, respectively), $t(39) = -4.61$, $p < .001$, $\eta^2 = .35$.

In order to assess the statistical impact of the 50–50 instruction, overall and in interaction with participant sample, we conducted two-way ANOVAs to compare students and investigators from the videotape conditions of Experiments 1 and 2. On global accuracy, a significant main effect indicated that students outperformed investigators (Ms = 53.6 and 45.3%, respectively), $F(1, 94) = 6.53$, $p < .01$, $\eta^2 = .07$. Although there were no significant main effects or interaction on hit responses, false alarms were significantly higher in the first Experiment than in the second (Ms = 62.0 and 49.4%, respectively), $F(1, 94) = 9.25$, $p < .001$, $\eta^2 = .09$, and among investigators than students (Ms = 63.3 and 48.0%, respectively), $F(1, 94) = 13.55$, $p < .001$, $\eta^2 = .13$. There was also a marginally significant interaction, which showed that the reduction in false alarms from the first experiment to the second was significant among investigators (Ms = 73.6 and 53.0%, respectively) but not among the students (Ms = 50.3 and 45.7%, respectively), $F(1, 94) = 3.70$, $p < .06$, $\eta^2 = .04$. Finally, confidence levels were higher in the first experiment than in the second (Ms = 6.91 and 6.34, respectively), $F(1, 94) = 6.79$, $p < .01$, $\eta^2 = .07$, and among investigators than students (Ms = 7.34 and 5.74, respectively), $F(1, 94) = 46.46$, $p < .001$, $\eta^2 = .33$.

Discussion

The primary aim of Experiment 2 was to neutralize the investigator response bias through a pre-task instruction that set the base rate for true and false

confessions at 50–50. This manipulation was successful, both in reducing the overall number of "true" judgments that had produced the response bias and in eliminating the differences between participant samples. The empirical question we raised was whether eliminating the response bias would improve performance, particularly within our sample of investigators. The results on performance measures were mixed. Compared to their counterparts in the video condition of Experiment 1, investigators in the video condition of Experiment 2 had a comparable hit rate but a lower false alarm rate, making them somewhat more accurate in their judgments. The problem is that while investigators in this study were not more accurate than students or chance performance, they were still overconfident in their judgments.

General Discussion

Analyses of recent DNA exonerations suggest that false confessions are implicated in more than 20% of all wrongful convictions. This problem occurs for two reasons: (1) people sometimes confess to crimes they did not commit, either voluntarily or through a process of interrogation, and (2) police investigators, district attorneys, judges, and juries seem unable to distinguish among true and false confessions, too often accepting the latter at face value. Archival and case studies illustrate the point. Looking at sixty proven and probable false confession cases, Leo and Ofshe (1998) discovered that 73% of defendants who were tried on the basis of these confessions were convicted.

Human beings and the criminal justice systems they create are imperfect. Defendants, police investigators, and witnesses make mistakes and lie, voluntarily or under pressure. Thankfully, there are safeguards in place to regulate the problems through adversarial mechanisms that press for corroboration, proof beyond a reasonable doubt, and post-conviction appellate review. In the case of confessions, the protection for people falsely accused rests on the commonsense assumption, held from the police station into the courtroom, that "I'd know a false confession if I saw one." This research challenges that assumption. In Experiment 1, police were not only less accurate than laypeople at judging whether confessions were true or false, they were also biased toward perceiving true confessions and overconfident despite a lack of accuracy. This pattern of results closely parallels studies of investigators asked to judge true and false denials (Meissner & Kassin, 2002; Garrido et al., 2004).

In addition to suggesting the fallacy of the belief that people can readily distinguish true and false confessions in the absence of other evidence, this research makes three new and important contributions. First, the results clarify the nature of the investigator response bias. Reanalyzing past studies from a signal detection framework, Meissner and Kassin (2002) discovered and then replicated a significant investigator response bias, a tendency for police to see deception in suspects. Using a standardized self-report instrument, Masip et al. (in press) found that police harbor a "generalized communicative suspicion" compared to others. But does this response disposition indicate a tendency to see deception or guilt? In forensic settings, lying and guilt are naturally conflated: innocent suspects state truthful alibis;

criminals lie in their denials. With confessions, however, in which "true" judgments indicate guilt and "false" judgments indicate innocence, we were able to test these competing explanations. The results were clear. Relative to students, investigators erred by accepting false confessions, not by rejecting true confessions. Hence, the bias is not to see lies per se, but to presume guilt. This result helps to explain another finding—that real life interviews were seen by police officers as more skillfully conducted when they elicited confessions than when they did not (Bull & Milne, 2004).

A second important contribution is in the finding that investigators continued to exhibit a performance pattern of low accuracy and high confidence even when this guilt bias was neutralized. Although this is the first study ever to assess judgments of true and false confessions, the results replicate a consistent finding that experience and training do not typically improve deception detection (Bull, 1989; Kassin & Fong, 1999; Porter et al., 2000; Vrij, 2000; Zuckerman et al., 1984) and that professionals perform only slightly better than civilians, if at all (DePaulo, 1994; DePaulo & Pfeifer, 1986; Ekman & O'Sullivan, 1991; Ekman et al., 1999; Garrido et al., 2004; Koehnken, 1987; Porter et al., 2000). We speculated that this poor performance was a by-product of the response bias we had previously discovered (Meissner & Kassin, 2002, 2004). Yet even when this bias was neutralized and the false alarm rate reduced in Experiment 2, this pattern persisted. In short, it appears that the performance problem among police stems from the use of nondiagnostic behavioral cues, such as gaze aversion (DePaulo et al., 2003) or a tendency to selectively focus on deception cues to the neglect of truth-telling cues (Garrido et al., 2004).

Third, this research showed that people are better judges of confessions when they listen to audiotapes of the statements than when they see complete audiovisual presentations. In Experiment 1, participants on average were 11.5% more accurate in the audiotape condition than in the videotape condition, and the change benefited both students (64.1% vs. 53.4%) and investigators (54.5% vs. 42.1%). This result is consistent with prior research indicating that people are better lie detectors when focused on content and auditory cues than on less diagnostic but distracting visual information (e.g., Anderson et al., 1999; DePaulo et al., 1982; Zuckerman, DePaulo, & Rosenthal, 1981). This finding raises an interesting policy question. In recent years, triggered in large part by DNA exonerations and concomitant discovery of false confessions, there has been discussion and movement in many states toward requiring the full electronic recording of all custodial interviews and interrogations (Drizin & Colgan, 2001; Kassin, 2004; Slobogin, 2003). This debate brings to light important logistical considerations, as suggested by the work of Lassiter and his colleagues on the impact of camera perspective on judges and juries (Lassiter, Geers, Munhall, Handley, & Beers, 2001; Lassiter, Geers, Handley, Weiland, & Munhall, 2002).

Based on the present finding that judgment accuracy was greater in the audiotape condition of Experiment 1 than the videotape condition, one might be tempted to draw from our results the recommendation that electronic recording be operationalized via audiotape recorders. However, any such conclusion would rest on the narrow view that the sole function of electronic recording is to improve the accuracy

of those who later assess the confessions. In fact, proposals for reform are motivated by other important goals as well, such as: to provide an objective record of all events that preceded the confession (e.g., whether *Miranda* warnings were administered and waived, whether threats or promises were made, whether the suspect was physically threatened, where the details contained in the confession came from), to deter police coercion and misconduct, to deter frivolous defense claims of coercion, to increase plea agreements, and to build trust in law enforcement. Research in nonforensic settings indicates that people become better lie detectors despite exposure to nondiagnostic visual cues if instructed to focus on the verbal and paralinguistic channels of communication (DePaulo et al., 1982). Although more research on this point is needed, it may be possible using appropriate focusing instructions to gain the benefits of a full videotaping requirement as well as increased judgment accuracy among police investigators, juries, and others.

These studies are limited in ways that might be addressed in future research. One limitation concerns our use of prison inmates as the population of confessors to be judged. We sought this population precisely because of their ability to offer true confessions to serious crimes actually committed as opposed to minor transgressions or mock crimes. Clearly, however, these participants seemed quite adept at lying, exhibiting little difficulty at the task of generating false confessions. For detection purposes, then, prison inmates may be a uniquely difficult target group to assess, which may be the reason they demonstrably outperform others at *detecting* deception (Hartwig, Granhag, Strömwall, & Andersson, 2004). A related concern is that although we checked with participating inmates to ensure that they had not ever committed the crime we had assigned for a false confession, we cannot discount the possibility that they inserted autobiographical truths into their fictitious stories, thus increasing the difficulty of the task. Indeed, when people were asked about the origins of their everyday lies, most said that they derived lies from actual experiences, altering critical details (Malone, Adams, Anderson, Ansfield, & DePaulo, 1997). Still another limitation concerns the motivational differences between our prison inmates and suspects who stand accused and whose performance bears consequence. Although our participants saw the task as challenging, they told their true and false stories in a relatively low-stakes situation, and did so in a matter of minutes, which can weaken deception cues and make the statements more difficult to judge (DePaulo et al., 2003).

The foregoing limitations suggest that the task confronting our participant observers was difficult, perhaps more so than in the interrogation room. It is important to note, however, that the accuracy rates observed in these studies are highly consistent with most past research, that the difficulty of the task does not account for the performance differences between students and investigators, and, from a metacognitive perspective, that investigators did not adjust their confidence levels downward in light of these paradigmatic limitations. One might argue that police investigators are trained to detect forensic high-stakes lies, but research has produced mixed results. Vrij and Mann (2001) found that police officers did not exceed chance level performance at judging the videotaped press conferences involving family members who pled for help in finding missing relatives that they had killed. Mann, Vrij, and Bull (2004) found that police did distinguish high-stakes truths and lies in

videotaped interviews, but these researchers tested participants on a per-statement basis rather than assess global judgments of guilt or innocence. They also did not independently vary the stakes or test a comparison group of laypersons. Hence, the elevated accuracy rates, relative to those obtained in prior studies, may reveal more about the task used than about the transparency of high-stake lies or accuracy of police officers.

One might also argue that investigators were limited by their ability merely to observe the confessions, not actively elicit them. However, research shows that judgments of truth and deception may be more accurate, and certainly are not less so, when made by observers than by conversational interactants (Buller, Strzyzewski, & Hunsaker, 1991; Hartwig, Granhag, Strömwall, & Vrij, 2004). In this regard, it is instructive that despite the possibility of this limitation, police participants not only completed the task but were highly confident in their judgments. In short, although we have raised possible concerns about the target persons and their statements, these concerns do not account for the significant differences between investigators and students (who, after all, judged the same confessions)—differences that closely parallel results from other laboratories when it comes to accuracy, confidence, and response bias. Indeed, students in the audiotape condition of Experiment 1 exhibited a 64% accuracy rate, which is precisely the rate Ekman and O'Sullivan (1991) obtained from secret service agents, their most proficient group of professionals.

One could reasonably argue that real life false confessions, which are commonly elicited through a process of interrogation, are more difficult to assess than those produced in this research. In most documented false confessions, as in the Central Park jogger case, the statements ultimately presented in court are highly scripted by investigators' theory of the case; they are rehearsed and repeated over hours of interrogation; and they often contain vivid details about the crime, the scene, and the victim that became known to suspects through secondhand sources (Kassin, 2002). Yet in this research, our prison inmates generated their false confessions immediately, spontaneously, without rehearsal, and without external assistance. This contrast raises an interesting empirical question for future research concerning the extent to which the interrogation techniques used to elicit full narrative confessions also increase the perception of credibility independent of the confessor's guilt or innocence.

The present results are provocative, but they represent a small first step in addressing the problem that false confessions are difficult to detect. At this point, additional research is needed to address a number of issues. On the stimulus side of the equation, an important next step is to present observers with *actual* taped confessions for which "ground truth" is known with certainty. This would address the question of whether interrogation-elicited statements are more or less difficult to assess than those spontaneously produced. Also important is to compare judgments of actual confessions that are made by observers who have or lack access to the full interrogation process. It is clear from much research that people are not adept at truth and lie detection from verbal and nonverbal demeanor. But this does not mean that the task is impossible. To the extent that diagnostic cues are inherent in the eliciting situation, performance may be enhanced in observers who see

not only the confession but the conditions under which it was elicited. Finally, on the respondent side, future research should seek to test not only police investigators and laypeople, but judges, prosecutors, and defense lawyers as well—actors within the legal system who would approach the task with different expectations and motives.

ACKNOWLEDGMENT

This research was supported by the Williams College Bronfman Science Center through funds awarded to the first author. We also want to thank administrators and staff members at Essex County Correctional Facility for their invaluable assistance in the recruitment and scheduling of the inmates who participated in this research.

REFERENCES

Anderson, D. E., DePaulo, B. M., Ansfield, M. E., Tickle, J. J., & Green, E. (1999). Beliefs about cues to deception: Mindless stereotypes or untapped wisdom? *Journal of Nonverbal Behavior*, *23,* 67–89.

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

Bull, R., & Milne, R. (2004). Attempts to improve the police interviewing of suspects. In G. D. Lassiter (Ed.), *Interrogations, confessions, and entrapment* (pp. 182–196). New York: 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.

Davis, D., & O'Donohue, W. (2003). The road to perdition: "Extreme influence" tactics in the interrogation room. In W. O'Donohue, P. Laws, & C. Hollin (Eds.), *Handbook of forensic psychology*. New York: Basic Books.

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

DePaulo, B. M., Lassiter, G. D., & Stone, J. I. (1982). Attentional determinants of success at detecting deception and truth. *Personality and Social Psychology Bulletin*, *8,* 273–279.

DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin*, *129,* 74–112.

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

Drizin, S. A., & Colgan, B. A. (2001). Let the cameras roll: Mandatory videotaping of interrogations is the solution to Illinois' problem of false confessions. *Loyola University Chicago Law Journal*, *32,* 337–424.

Drizin, S. A., & Leo, R. A. (2004). The problem of false confessions in the post-DNA world. *North Carolina Law Review*, *82,* 891–1007.

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.

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

Garrido, E., Masip, J., & Herrero, C. (2004). Police officers' credibility judgments: Accuracy and estimated ability. *International Journal of Psychology*, *39,* 254–275.

Granhag, P.-A., & Strömwall, L. (Eds.) (2004). *The detection of deception in forensic contexts*. Cambridge: Cambridge University Press.

Gudjonsson, G. H. (1992). *The psychology of interrogations, confessions, and testimony*. London: Wiley.

Gudjonsson, G. H. (2003). *The psychology of interrogations and confessions: A handbook*. West Sussex, UK: Wiley.

Hartwig, M., Granhag, P. A., Strömwall, L. A., & Andersson, L. O. (2004). Suspicious minds: Criminals' ability to detect deception. *Psychology, Crime, and Law*, *10,* 83–95.

Hartwig, M., Granhag, P. A., Strömwall, L. A., & Vrij, A. (2004). Police officers' lie detection accuracy: Interrogating freely vs. observing video. *Police Quarterly*, *7,* 429–456.
Hilgendorf, E. L., & Irving, M. (1981). A decision-making model of confessions. In M. Lloyd-Bostock (Ed.), *Psychology in legal contexts: Applications and limitations* (pp. 67–84). London: MacMillan.
Inbau, F. E., Reid, J. E., Buckley, J. P., & Jayne, B. C. (2001). *Criminal interrogation and confessions* (4th ed.). Gaithersburg, MD: Aspen.
Kassin, S. M. (1997). The psychology of confession evidence. *American Psychologist*, *52,* 221–233.
Kassin, S. M. (2002). False confessions and the jogger case. *The New York Times*, November 1, 2002, p. A-31.
Kassin, S. M. (2004). Videotape police interrogations. *The Boston Globe, OP-ED*, April 26, 2004, p. A-13.
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., & Neumann, K. (1997). On the power of confession evidence: An experimental test of the "fundamental difference" hypothesis. *Law and Human Behavior*, *21,* 469–484.
Kassin, S. M., & Sukel, H. (1997). Coerced confessions and the jury: An experimental test of the "harmless error" rule. *Law and Human Behavior*, *21,* 27–46.
Kassin, S. M., & Wrightsman, L. S. (1980). Prior confessions and mock juror verdicts. *Journal of Applied Social Psychology*, *10,* 133–146.
Kassin, S. M., & Wrightsman, L. S. (1985). Confession evidence. In S. M. Kassin & L. S. Wrightsman (Eds.), *The psychology of evidence and trial procedure* (pp. 67–94). Beverly Hills, CA: Sage.
Koehnken, G. (1987). Training police officers to detect deceptive eyewitness statements: Does it work? *Social Behavior*, *2,* 1–17.
Lassiter, G. D. (Ed.) (2004). *Interrogations, confessions, and entrapment*. New York: Kluwer Academic Publishers.
Lassiter, G. D., Geers, A., Handley, I., Weiland, P., & Munhall, P. (2002). Videotaped confessions and interrogations: A simple change in camera perspective alters verdicts in simulated trials. *Journal of Applied Psychology*, *87,* 867–874.
Lassiter, G. D., Geers, A. L., Munhall, P. J., Handley, I. M., & Beers, M. J. (2001). Videotaped confessions: Is guilt in the eye of the camera? *Advances in Experimental Social Psychology*, *33,* 189–254.
Leo, R. A. (1996). Inside the interrogation room. *The Journal of Criminal Law and Criminology*, *86,* 266–303.
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,* 429–496.
Malone, B. E., Adams, R. B., Anderson, D. E., Ansfield, M., & DePaulo, B. M. (1997). *Strategies of deception and their correlates over the course of friendship*. Poster presented at the annual meeting of the American Psychological Society, Washington, DC.
Mann, S., Vrij, A., & Bull, R. (2004). Detecting true lies: Police officers' ability to detect suspects' lies. *Journal of Applied Psychology*, *89,* 137–149.
Masip, J., Alonso, H., Garrido, E., & Anton, C. (in press). Generalized Communicative Suspicion (GCS) among police officers: Accounting for the investigator bias effect. *Journal of Applied Social Psychology.*
Meissner, C. A., & Kassin, S. M. (2002). "He's guilty!": Investigator bias in judgments of truth and deception. *Law and Human Behavior*, *26,* 469–480.
Meissner, C. A., & Kassin, S. M. (2004). "You're guilty, so just confess!" Cognitive and behavioral confirmation biases in the interrogation room. In D. Lassiter (Ed.), *Interrogations, confessions, and entrapment*. New York: Kluwer Academic/Plenum Press.
Memon, A., Vrij, A., & Bull, R. (2003). *Psychology and law: Truthfulness, accuracy and credibility*. London: Jossey-Bass.
Morgenthau, R. (2002). New York v. Wise, Richardson, McCray, Salaam, & Santana: Affirmation in Response to Motion to Vacate Judgment of Conviction. Indictment No. 4762/89.
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.
Redlich, A. D., & Goodman, G. S. (2003). Taking responsibility for an act not committed: The influence of age and suggestibility. *Law and Human Behavior*, *27,* 141–156.
Saulny, S. (2002). Why confess to what you didn't do? *The New York Times*, December 8, 2002, Section 4.
Scheck, B., Neufeld, P., & Dwyer, J. (2000). *Actual innocence*. New York: Doubleday.
Slobogin, C. (2003). Toward taping. *Ohio State Journal of Criminal Law*, *1,* 309–322.
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. (2000). *Detecting lies and deceit: The psychology of lying and the implications for professional practice*. London: Wiley.

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.

Wrightsman, L. S., & Kassin, S. M. (1993). *Confessions in the courtroom*. Newbury Park, CA: Sage.

Zimbardo, P. G. (1967, June). The psychology of police confessions. *Psychology Today*, *1,* 17–20, 25–27.

Zuckerman, M., DePaulo, B. M., & Rosenthal, R. (1981). Verbal and nonverbal communication of deception. *Advances in Experimental Social Psychology*, *14,* 1–59.

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