Law Hum Behav (2011) 35:327–337
DOI 10.1007/s10979-010-9244-2

ORIGINAL ARTICLE

Inside Interrogation: The Lie, The Bluff, and False Confessions

Jennifer T. Perillo · Saul M. Kassin

Published online: 24 August 2010

Abstract Using a less deceptive variant of the false evidence ploy, interrogators often use the bluff tactic, whereby they pretend to have evidence to be tested without further claiming that it necessarily implicates the suspect. Three experiments were conducted to assess the impact of the bluff on confession rates. Using the Kassin and Kiechel (Psychol Sci 7:125–128, 1996) computer crash paradigm, Experiment 1 indicated that bluffing increases false confessions comparable to the effect produced by the presentation of false evidence. Experiment 2 replicated the bluff effect and provided self-reports indicating that innocent participants saw the bluff as a promise of future exoneration which, paradoxically, made it easier to confess. Using a variant of the Russano et al. (Psychol Sci 16:481–486, 2005) cheating paradigm, Experiment 3 replicated the bluff effect on innocent suspects once again, though a ceiling effect was obtained in the guilty condition. Results suggest that the phenomenology of innocence can lead innocents to confess even in response to relatively benign interrogation tactics.

Keywords Interrogation · Innocence · False confessions

Despite its historical status as something of a gold standard, confession evidence has been implicated in numerous wrongful convictions—including 20–25% of DNA exoneration cases in the United States (Garrett, 2008; http://www.innocenceproject.org/). On the question of why people would confess to crimes they did not commit, research has identified two sets of risk factors: (1) Dispositional vulnerabilities inherent in the suspect, such as youth, intellectual impairments, mental illness, and personality traits that foster compliance and suggestibility (see Gudjonsson, 2003; Kassin & Gudjonsson, 2004) and (2) situational pressures inherent in the conditions of custody and interrogation, such as excessive time, the presentation of false incriminating evidence, and the use of minimization themes that imply leniency (Kassin, 2007; Leo, 2008; for recent reviews, see Kassin et al., 2010).

Many police investigators are trained in use of the Reid technique, a multistep approach to interrogation that is designed to increase the anxiety associated with denial while reducing the anxiety associated with confession (Inbau, Buckley, Reid, & Jayne, 2001). Perhaps, the most controversial tactic permissible within this approach involves the false evidence ploy by which interrogators bolster an accusation by presenting the suspect with supposedly incontrovertible evidence of his or her guilt (e.g., a fingerprint, blood or hair sample, eyewitness identification, or failed polygraph)—even if that evidence does not exist. In the United States, this type of deception is permitted by law (*Frazier v. Cupp*, 1969), recommended under certain circumstances (Inbau et al., 2001), and occasionally used by police (Kassin et al., 2007; Leo, 1996).

Drawing on the distinction between coerced-compliant and internalized false confessions (Kassin & Wrightsman, 1985), there are two mechanisms by which innocent suspects might confess when confronted with false evidence. First, research has shown that people in general confess as an act of social compliance when they feel trapped by the apparent strength of the evidence against them and perceive no other means of escape (Gudjonsson & Sigurdsson, 1999; Moston, Stephenson, & Williamson, 1992). Second, research has shown that false evidence, a strong form of

J. T. Perillo · S. M. Kassin (✉)
Department of Psychology, John Jay College of Criminal Justice, 445 West 59 Street, New York, NY 10019, USA
e-mail: skassin@jjay.cuny.edu

misinformation, can create confusion and lead people to doubt their own beliefs, at times internalizing guilt and confabulating memories for crimes they did not commit (Kassin, 2007; for a more general review of the misinformation effect, see Loftus, 2005).

Both of these effects have been demonstrated in laboratory experiments. In the first such study, Kassin and Kiechel (1996) developed a paradigm by which participants thought they were engaged in a reaction time study that involved typing letters on a keyboard that were read aloud by a confederate. During the task the computer crashed, at which point the experimenter falsely accused participants of pressing the ALT key they were instructed to avoid and asked them to sign a handwritten confession. In this initial experiment, the presentation of false evidence was varied by having the confederate, serving as a witness, report that she did or did not see the participant hit the forbidden key. Across conditions, this manipulation nearly doubled the number of signed false confessions, from 48 to 94%. As measured by participants' private admissions to a second confederate, the manipulation also increased the number who internalized guilt, from 12 to 55%. Indicating that deception increases the risk of compliant and internalized false confessions, these results have been replicated in several studies (e.g., Horselenberg, Merkelbach, & Josephs, 2003; Redlich & Goodman, 2003; for a replication within a different paradigm, see Nash & Wade, 2009) even among informants who are pressured to report on a confession allegedly made by another person (Swanner, Beike, & Cole, 2010).

Anecdotal evidence from actual cases shows this false evidence effect is not a mere laboratory phenomenon. The 1989 case of 17-year-old Marty Tankleff, who was wrongly convicted for the murder of his parents, illustrates the point. During a five-plus hour interrogation, the lead detective outright lied to Tankleff about the evidence—e.g., claiming that his hair was found in his mother's grasp and that his father, who was in a coma, regained consciousness and identified his son as the attacker. By citing the most trusted source in his life, police led Tankleff to wonder if he had blacked out and murdered his parents, ultimately leading him to question his own innocence. On the basis of a confession he gave but quickly retracted, Tankleff was convicted. Nineteen years later, his conviction was overturned and all charges were dismissed (Firstman & Salpeter, 2008).

In addition to the dispositional and situational factors that can increase one's vulnerability to false confession, Kassin (2005) suggested the paradoxical hypothesis that innocence itself may put innocent people at risk to confess. Research shows that innocent people who stand accused believe that their innocence will become apparent to others, a naive belief that leads them to waive their *Miranda* rights to silence and to an attorney (Kassin & Norwick, 2004), and their right to a lineup (Holland, Kassin, & Wells, 2005), behave in ways that are forthcoming and cooperative in their interviews with police (Hartwig, Granhag, Strömwall, & Kronkvist, 2006; Hartwig, Granhag, Strömwall, & Vrij, 2005), and exhibit less physiological arousal on critical items of a concealed information test even when they are pre-informed about the crime (Zvi, Nachson, & Elaad, 2010). Ironically, laboratory research suggests that mock interrogators who presume a suspect guilty conduct particularly pressure-filled interrogations when paired with mock suspects who are truly innocent and adamant in their denials (Kassin, Goldstein, & Savitsky, 2003).

If innocence as a state of mind leads people to trust the criminal justice system, then interrogation techniques designed to buffer innocent suspects may not have the intended effect. In particular, consider the bluff technique, a common and less deceptive alternative to the false evidence ploy. In using the bluff, interrogators pretend to have evidence without additionally asserting that this evidence necessarily implicates the suspect (e.g., stating that witnesses were present and will be interviewed; or that blood, hair, or other biological evidence was collected and sent to a laboratory for testing). Some interrogation manuals specifically recommend using case files, dossiers, and other visual aids as a means of bolstering the bluff (e.g., Inbau et al., 2001; Zulawski & Wicklander, 2002).

In principle, the bluff should produce diagnostic outcomes by threatening the actual perpetrator with certain detection, increasing the incentive to cooperate and confess (as with the presentation of strong evidence discussed earlier; see Moston et al., 1992). Yet innocent suspects should not feel similarly threatened by the future evidence implied by the bluff. Knowing that they have left no trail of incriminating evidence, they should not fear being implicated and, hence, should not confess as a way to cut anticipated losses. Indeed, as Inbau et al. (2001) put it: "Merely introducing fictitious evidence during an interrogation would not cause an innocent person to confess. It is absurd to believe that a suspect who knows he did not commit a crime would place greater weight and credibility on alleged evidence than his own knowledge of his innocence" (p. 429).

The hypothesis that the bluff technique should elicit confessions from perpetrators but not from innocents makes intuitive sense. Based on actual case anecdotes, however, Kassin (2005) suggested the contrary hypothesis that to an innocent suspect under interrogation, the "threat" of proof implied by the bluff represents a "promise" of future exoneration, paradoxically making it easier to confess. In a case that illustrates how the phenomenology of innocence can wreck havoc on a suspect's decision making, Todd Johnson, who was ultimately acquitted, had

confessed to his wife's murder after 19 h of interrogation when police said that they found blood in his car to be sent to a laboratory for DNA testing. Knowing that the blood could not be his wife's, this defendant explained later that he confessed because he was exhausted and knew that the test results would show his innocence (*Missouri v. Johnson*, 2001). This logic was also revealed in the case of Jeffrey Deskovic. During a 6-h interrogation, police asserted that they had collected DNA at the rape and murder scene for testing. At that point, despite—or because of—his innocence, Deskovic confessed and was later convicted. After his exoneration, he explained why he confessed: "Believing in the criminal justice system and being fearful for myself, I told them what they wanted to hear." Knowing that the DNA testing would show his innocence, he said, "I thought it was all going to be O.K. in the end" (Santos, 2006). Interestingly, there was DNA evidence in this case and the test did exclude Deskovic. He was tried anyway, however, and convicted by a jury. Sixteen years later, he was released when the DNA was matched to the actual perpetrator.

These stories raise two important points about the bluff as an interrogation tool. First, it need not make suspects feel pressured or trapped, or alter their beliefs and memories, to elicit a confession. Rather, innocent people may "voluntarily" decide to confess out of a misplaced confidence that their admission will later be disproved. Second, it should not matter whether the police cite future evidence as a tactical bluff or as a truthful matter of fact. To the extent that an innocent suspect harbors the hope of future exoneration, the net effect may be the same: to increase the risk of confession.

The research reported in this article presents the first empirical examination of the bluff tactic on confessions. Using the computer crash paradigm described earlier, a first study was designed with three goals in mind: (1) to investigate for the first time the effect of the bluff tactic on compliant and internalized false confessions, (2) to compare the bluff to the false evidence ploy, and (3) to determine whether the presence of a witness who affirms the participant's denial protects that participant against false confessions.

Experiment 1

Method

Participants and Design. Seventy-nine introductory psychology students (51 female, 28 male; mean age = 20.2 years) at a large urban university participated in exchange for course credit. Each participant was randomly assigned to one of five groups: (1) False witness evidence, (2) the bluff, (3) false witness evidence and bluff combined, (4) no-tactics control, and (5) witness-affirmed innocence control. Data from seven participants were excluded from analyses because they expressed suspicion to the confederate after the experimenter departed; an eighth was excluded because he confessed before the experimental manipulation was performed. The final sample thus consisted of 71 participants (*n*'s of 13–15 per cell).

Procedure. Participants were recruited for a reaction time study to be conducted in pairs (in actuality, each session involved one participant and one female confederate). After filling out a cover story questionnaire concerning spatial awareness, keyboard experience, and reflexes, the pair was taken into the laboratory and seated next to each other, across the table from a female experimenter. At that point, the experimenter explained that one participant would read a series of letters at a fixed pace and the other would type those letters on a desktop computer keyboard as quickly and as accurately as possible using alternating forefingers. Through a rigged "random" draw, the participant was always assigned to the typist role first; the confederate was the assigned reader.

Before the start of the task, the experimenter told participants that there was a glitch in the program. Specifically, she warned against pressing the ALT key, which would cause the computer to crash and the loss of data that were stored. After ensuring that the participant (and confederate) understood this rule, she directed them to start the task. Corresponding to the slow-pace condition of Kassin and Kiechel (1996), the confederate read aloud a list of stimulus letters at a pace of 42 letters per minute. Approximately 60 s into the session, the computer gave off a beeping "Windows Error" alarm. Startled, the experimenter accused the participant of hitting the forbidden ALT key. She then grabbed the keyboard and tried to shut down the program, forcing the computer to reboot. While waiting, the experimenter turned to the participant and asked, "Did you hit the ALT key?" All participants (except one who was excluded for having actually hit the ALT key and confessing as such) initially—and, in some cases, vehemently—denied hitting the key. After rebooting, the experimenter became visibly distressed as she apparently failed to locate the program on the hard drive and noted that there was no backup. Once again, in an exasperated tone, she turned to the participant and asked, "Are you sure you didn't hit the ALT key?"

This question was followed by the first experimental manipulation, the presentation of *false incriminating evidence*. After the participant denied hitting the key for a second time, the experimenter turned to the confederate and asked, "Did you see anything?" In the false evidence

condition, the confederate said that she saw the participant hit the ALT key with the side of his or her finger. In the no-tactics control condition, the same confederate said she was too busy reading to see anything.

Secondly, we varied the presentation of the *bluff technique*. For participants in the bluff condition, the experimenter stated that the computer they had typed on was connected to a server located in another room and that all keystrokes from experimental sessions had been recorded, making it possible to check on whether the ALT key had been hit. The experimenter went onto explain, however, that because the server was password-protected by her professor, she was not able to check it until she could locate him. In the *no-bluff* condition, the experimenter said nothing about recorded keystrokes or a computer server in another room.

Finally, we sought to determine whether the presentation of *affirmative innocence evidence*—as seen in cases in which suspects have strong alibis, have passed a polygraph, or were excluded by DNA or other forensic tests—would help buffer participants from the pressure to sign a false confession. To test this hypothesis, we ran a second baseline control group (i.e., neither false evidence nor the bluff was used), what might be called the *affirmed innocence* control, in which the confederate-witness answered the experimenter's query by reporting that the participant's hands were nowhere near the ALT key.

After the experimental manipulations, the experimenter, still agitated, ripped a sheet of paper from a notepad and handwrote the following statement: "I hit the ALT key I wasn't supposed to press and caused the program to crash. Data was lost. Session was terminated without credit." The experimenter then presented the statement to the participant with a request to read and sign it. If the participant refused, the request was repeated a second time. Whether or not the participant signed the statement provided a behavioral measure of *compliant* false confessions.

Regardless of the participant's decision, the experimenter then stopped the session, administered post-session questionnaires, and left the room. These questionnaires were meant to reinforce the cover story and give the confederate an opportunity to probe the participant for suspicion privately, away from the experimenter. The confederate's questionnaire concerned the reading task; the participant's concerned the typing task (e.g., it asked participants to rate how difficult they found the task, how clear the confederate was at reading, and how accurately they performed). While filling out the questionnaire, the confederate probed for suspicion and for evidence that the participant had exhibited *internalization,* or belief in his or her culpability. The confederate first asked, "Why did you (not) sign the confession?" After the participant responded, the confederate then asked, "What do you think is going to happen?" To assess internalization, the participant's responses to the experimenter's probes were audiotape recorded and later coded for evidence of belief in guilt. As in Kassin and Kiechel (1996), a stringent standard for internalization was set whereby participants had to indicate without ambiguity the belief that they had hit the ALT key (e.g., "I hit the key," "I broke the computer," "It was my fault"). Statements accompanied by an uncertainty (e.g., "I think" or "I may have") were not scored as instances of internalization. All responses were coded by two raters whose agreement rate was 100%.

Once the questionnaires were completed, the experimenter further queried the participant for suspicion. She then fully debriefed the participant, explaining the purpose of the study as well as the need for deception.

Results

Overall, 43 of 71 (60.6%) participants confessed to pressing the forbidden ALT key, and 7 of 71 (9.9%) internalized a belief in their culpability. The total percentage of compliant false confessions was comparable to the 69% rate reported by Kassin and Kiechel (1996); the rate of internalizations was lower than their 28% rate. As will be discussed shortly, the latter difference is consistent with the proposed mechanism by which the bluff leads innocents to confess without a change in belief.

On the question of whether participants signed the confession, there was a significant difference across conditions, χ^2 (4, $N = 71$) = 18.47, $p < .001$, Cramer's $V = 0.51$, 95% CI .03 to .36. Replicating past research, specific comparisons showed that the presentation of false evidence, on its own, significantly increased the confession rate relative to the no-tactic control group (79 vs. 27%), χ^2 (1, $N = 29$) = 7.813, $p < .01$, Cramer's $V = .52$, 95% CI .01 to .23. Importantly, and contradicting the theory underlying use of the more nuanced bluff technique, the false confession rate was also significantly increased in the bluff-only group (87 vs. 27%), χ^2 (1, $N = 30$) = 11.00, $p < .01$, Cramer's $V = .61$, 95% CI .02 to .28. As seen in Table 1, the false confession rate was as high in the

Table 1 Percentages of compliant and internalized false confessions in Study 1

Condition	Confessions (%)	Internalization (%)
Baseline control	26.67[a]	00.00
Innocence control	35.71[a]	00.00
False evidence	78.57[b]	14.29
Bluff	86.67[b]	06.67
Both	76.92[b]	30.77

Note: Percentages not sharing a common superscript differ at $p < .05$

bluff-only group as it was when the false evidence ploy was presented alone or when the two tactics were presented in tandem. Interestingly, the introduction of an innocence-affirming witness did not buffer participants from the accusation and pressure of the situation (the false confessions rates were 36 vs. 27%, respectively, in the innocence and baseline control groups).

With regard to the measure of internalization, chi-square analyses indicated that there was a significant between-groups difference, χ^2 (4, $N = 71$) = 10.05, $p < .05$, Cramer's $V = 0.38$, 95% CI .00 to .22. Specific comparisons, however, did not yield significant increases in either the false evidence or bluff-only conditions—only a marginally significant increase when both tactics were used, Yates-corrected χ^2 (1, $N = 28$) = 3.16, $p < .10$, 95% CI .00 to .14.

Experiment 2

Using the computer crash paradigm, the first experiment provided strong support for the hypothesis that the bluff technique—a quasi-deceptive version of the false evidence ploy—can induce innocent people to confess. Led to believe that their keystrokes would be retrieved, participants became more likely to sign a confession, not less so, exhibiting an increase by 60% over the no-tactics control group. This effect was as potent as that produced by the presentation of false evidence. Contrary to the wisdom that innocents would not react to the potentially threatening implication of the bluff, this result suggested that mere bluffing can put innocent suspects at risk. As indicated by the low rates of internalization, it is clear that the bluff elicits false confession strictly as an act of compliance without altering beliefs in culpability, as in the false evidence ploy.

Although the results are consistent with our hypothesis concerning the phenomenology of innocence, no direct self-report evidence was obtained for the proposed mechanism. A second experiment was therefore conducted both to replicate the bluff effect and to collect post-decision self-reports to determine if participants would articulate that they confessed as a result of the belief in a promise of future exoneration. Social psychologists have long known that people are often not adept at identifying the causes of their own behavior (Wilson & Nisbett, 1978). In this paradigm, however, as in the retrospective anecdotal reports of wrongfully convicted confessors (Kassin, 2005), we predicted that our participants—precisely because of their known innocence—would consciously choose to confess because the expectation for exoneration would lead them to underweight the net cost of confession versus denial.

Method

Participants and Design. Forty-four introductory psychology students (34 female, 10 male) at a large urban university were recruited in exchange for course credit. Each participant was randomly assigned to either the bluff or the baseline control condition. Data from six participants were excluded from analyses because they expressed suspicion to the confederate after the experimenter departed or had already heard about the study. The final sample thus consisted of 38 participants ($n = 19$ per group).

Procedure. The procedures were identical to those of Experiment 1 except that students were assigned to only the bluff or the control condition. In addition, they were queried after being told that the experiment was rigged to elicit their self-reported reasons for confession or denial. After probing for suspicion, as in the first study, the experimenter informed participants that the computer did not crash and that the true purpose of the study was to investigate people's decision making with regard to whether to sign the requested statement. The experimenter then asked a series of open-ended questions in an interview that was covertly audiotaped for later analysis. In this orally conducted interview, the following questions were asked: (1) "Can you tell us what you were thinking about when you were deciding whether or not to sign the statement?" and (2) "Did anything I say play a role in your decision to sign or not sign?" Control group participants were terminated at this point, as were those in the bluff group who cited the bluff as a basis for their decision (e.g., "[I was expecting] that you were going to go over there and check and see that I didn't touch it, and be like, alright."). Those in the bluff condition who did not cite the manipulation were asked two additional questions: (3) "What about the server that I mentioned had recorded your keystrokes in the other room—Did that play a role in your decision to sign?" and (4) "Did the recorded keystrokes make it easier or harder for you to sign the statement?" Once the interview was completed, the experimenter revealed the true purpose of the study, obtained signed permissions to use the audiotaped data, and fully debriefed the participants.

Results

Overall, 22 of 38 (57.9%) participants confessed to hitting the forbidden ALT key; none internalized a belief in their culpability. Replicating the main result of first experiment, the bluff significantly increased the false confession compliance rate, from 47 to 74%, χ^2 (1, $N = 38$) = 3.89, $p < .05$, Cramer's $V = .32$, 95% CI .00 to .15.

With regard to self-reported explanations for confession and denial, four participants had to be excluded from analyses because of audiotape failures. The remaining 34 interviews were coded by two independent raters who were, at the outset, blind to condition ($\kappa = .811$). Raters also coded for participants' self-reported and spontaneous expressions of certainty in their own innocence in response to probes by the confederate and experimenter (e.g., "What happened?"; "What were you thinking about when deciding whether or not to sign?"). Self-reported certainty ratings were collected to determine if participants who had confessed said they did so because the bluff offered a promise of future exoneration—not because they had come to question their innocence, as is often found when false evidence is presented. These self-reports were coded on a five-point scale ranging from completely certain of innocence (e.g., "I'm sure I did not hit the ALT key") to completely certain of guilt (e.g., "I know I hit the ALT key") to provide a measure of guilt certainty (inter-rater $r = .670$, $p < .001$). Disagreements were resolved by an independent third coder.

Overall, 94% of participants expressed some degree of certainty in their own innocence: 24 (73%) were completely certain; 7 (21%) were somewhat to mostly certain; 2 (6%) said they were somewhat certain of their guilt. Despite the fact that most participants knew they were innocent, however, a majority agreed to confess. This result further supports the hypothesis that participants in the bluff condition signed the confession not as a result of internalization but as an act of compliance stemming from the expectation that ultimately they would be exonerated despite the confession.

Self-reports provided additional evidence of this proposed mechanism. Specifically, 75% of those who confessed in the bluff condition explicitly cited the bluff as a reason for that decision (e.g., "I think that [the program] made it easier because you said that you would come back and tell me. So I was like, I'm pretty sure that I didn't press it. So I thought it would have been, I wouldn't be guilty. So I would have been ok."). Interestingly, three participants in the bluff condition who would not confess also cited the bluff as a reason for their decision (e.g., "You know, I was willing to sit here and wait [for the program to be checked] because I knew I didn't hit the key."). Obviously, no participants cited the bluff in the no-bluff control group.

Across conditions, 55% of participants who confessed cited interrogation pressure as a reason for confessing—compared to only 21.4% of those who did not confess, $\chi^2 (1, N = 34) = 3.83$, $p = .05$, Cramer's $V = .336$, 95% CI .00 to .15. It is not surprising that "pressure" was more often cited as a basis for confession than non-confession. Indeed, the psychological approach to police interrogation is specifically designed to increase a suspect's anxiety associated with denial, while reducing the anxiety associated with confession (Inbau et al., 2001). Perpetrators should be motivated to confess by a desire to escape the pressure of interrogation and cut anticipated losses. However, it appears that same pressure may cause innocent suspects to confess as well—and for them the bluff can make that decision easier by lessening the anticipated negative repercussions of confession.

Finally, note that in addition to citing the bluff and interrogation pressure as reasons for confession, many participants cited a disparate collection of miscellaneous "other" reasons—namely, the desire to leave; sympathy for the experimenter; the low stakes of the confession; an unwillingness to accept guilt, lose credit, or be held responsible in any way; and a failure to read the statement.

Experiment 3

Using the computer crash paradigm, Experiments 1 and 2 provided strong support for the predicted but paradoxical effect of the bluff on false confessions. In both the studies, a significant majority of participants confessed in response to this tactic; in the second study, three quarters of those who confessed in the bluff condition also reported that the bluff—and their expectation of future exoneration—played a role in their decision. While a small number of participant self-reports supported the opposing theory that the bluff provides innocent people with a reason *not* to confess, a vast majority confessed despite innocence on the basis of the promise of future exoneration.

One could argue that even if the bluff technique increases the risk that innocent suspects confess, that risk is outweighed by a substantial boost in the number of perpetrators who also are induced to confess. Clearly, the surgical objective of interrogation is to secure confessions from suspects who are guilty but not from those who are innocent. The process should thus be built to produce outcomes that are diagnostic of guilt and innocence, as measured by the observed ratio of true to false confessions. To assess the effect of the bluff on *diagnosticity*, a third experiment was conducted both to replicate the effect in a different experimental paradigm and to measure the true and false confession rates, respectively, when both perpetrators and innocents are tested.

To achieve these objectives, we developed a variant of the cheating paradigm that was first used to induce laboratory confessions to a willful act and to test the reactions of both guilty and innocent participants (Russano, Meissner, Narchet, & Kassin, 2005). In this procedure, participants were recruited to take part in a study of individual and joint problem-solving. They were instructed that they would be asked to answer an alternating series of

questions—some alone; others jointly with a fellow participant—who was actually a confederate. In the guilty condition, the confederate pretended to struggle on an individual problem, whereupon she asked for help, thus coaxing participants into violating the prohibition against collaborating during individual trials. In the innocent condition, the confederate did not ask for help, so participants were not coaxed into breaking the rule. At the conclusion of each session, the experimenter reported the "problem" to the pair that too many of their individual answers matched perfectly, separated them, and accused the participant of cheating. After noting the seriousness of the offense and citing the college's honor code, the experimenter sought to obtain a confession.

There are two benefits of the cheating paradigm. First, it addresses a limitation of the computer crash studies—namely, that they elicit confessions to an act of mere negligence, not one of willfulness, intent, and consequence (e.g., for the participant's academic record, as implied by the experimenter's reference to the university's honor code). Second, it allows us to investigate the extent to which the bluff, or some other tactic, increases the true confession rate among actual perpetrators. By independently varying guilt and innocence, we can test the effects of the bluff on both groups and thereby assess the diagnosticity of the confessions that are produced (for a fuller discussion of this paradigm and its benefits, see Meissner, Russano, & Narchet, 2010).

Method

Participants and Design. Seventy-two introductory psychology students at a large urban university were recruited in exchange for course credit (52 female, 20 male). Each participant was randomly assigned to one of four cells produced by a 2 (guilty vs. innocent) × 2 (bluff vs. no bluff) between-subjects factorial design. Data from seven participants were excluded because they expressed suspicion to the confederate. Another six were excluded because they failed to conform to the experimental manipulations (three guilty condition participants refused to help the confederate, two innocent condition participants blurted out their answers without prompting, and one left early before the experimenter sought a confession). The final sample thus consisted of 60 students (*n*'s of 14–16 per cell).

Procedure. Participants were recruited for a "social intelligence" study and asked to fill out an alternating series of ten questionnaires to investigate whether dyads exhibit higher rates of social intelligence. The questionnaires were designed to mimic the game show *Family Feud*, where contestants are given a category (e.g., People Who Get Tips; Things You Rent) and asked to list the four most popular survey answers within that category. Each questionnaire consisted of four content areas. Participants were told not to collaborate on the individual questionnaires or the research results would be invalid. Using this modified version of the Russano et al. (2005) cheating paradigm, all participants were paired with a confederate who manipulated guilt and innocence by inducing cheating in some sessions but not others. Sessions took approximately 30 min to complete.

In all sessions, the experimenter started the pair with an individual question series. After 3 min, she returned, collected the questionnaires, and administered a group question series to complete. After 3 min, the experimenter returned again, collected the group questionnaire, and handed them the second individual series. In the guilty condition, the confederate pretended to struggle at this point. Appearing panicked, the confederate turned to the participant and asked, "I'm totally blanking on answers for these. What did you get?" If the participants answered, the confederate wrote down the answers given. If the participant did not respond, the confederate repeated the request. In the innocent condition, the confederate completed the second individual question series in silence.

After the second individual trial was complete, the experimenter returned, collected the questionnaires, and administered the second group series. Two minutes into this round, however, she reentered, stated there was a problem, and asked the participant (who was always seated closest to the door) to follow her. In a separate room, the experimenter (who was blind to the participants' actual guilt or innocence) then told the participant, "It appears that the two of you collaborated on the last individual question series. I was just reviewing your answers, and, statistically, there is no way the two of you could have gotten so many of the same answers without working together." The experimenter cited the university's honor code and asked the participant if he or she had collaborated on the individual question series.

In half of the conditions, if the participant denied cheating, which they all initially did, the experimenter presented the second experimental manipulation, the *bluff*. In the bluff condition, she noted that a hidden camera stationed in an adjacent room with a one-way mirror had taped the entire session. She went on to explain, however, that the session file downloads directly onto a hard drive that could only be accessed by a video technician who was not scheduled to be in for several hours. In the control condition, the experimenter offered no potential proof to substantiate her belief that the participants had cheated.

Following this manipulation, the experimenter handwrote a confession and asked the participant to read and sign it: "I admit that I shared information with the other

participant on an individual problem series. Session was terminated without credit." If the participant refused, the request was reiterated along with a repeated reference to the recording in the bluff condition ("Look, if we see that you didn't do it, obviously you won't get in trouble. But if we see that you did it, you're going to lose your credit either way, so there's no point in wasting time"). After the participant's decision to sign or not sign the statement, the experimenter returned the participant to the original room with the confederate while she left, ostensibly to find her professor. During this time, the confederate probed for suspicion and internalization in the same manner as was done in Experiments 1 and 2 (the confederate first asked what happened and what the participant did before asking why). To allow for further coding, this interaction was covertly audiotaped as was the session with the experimenter that followed.

Afterward, the experimenter returned to obtain participants' self-reported reasons for confessing. The first two probes were the same used in Experiment 2. At that point all no-bluff participants were terminated, as were all bluff participants who had already cited the bluff as a factor in their decision making (e.g., "I was thinking that if you have cameras here, you can check it and it would show that we didn't cheat."). If participants in the bluff condition did not refer to the manipulation, the experimenter asked a third question: "What about the secret camera I told you about. Did that play a role in your decision?" Fourth, the experimenter asked, "Did the secret videotape make it easier or harder for you to (not) sign the statement?" After all questions were answered, the experimenter revealed the true purpose of the study, obtained permission to use the covertly audiotaped data, and fully debriefed the participant.

Results

Overall, 34 of 59 (57.6%) participants confessed to collaborating with the confederate on an individual task. To assess the effect of the bluff on perpetrators and innocents, a 2 (guilt vs. innocence) × 2 (bluff vs. no bluff control) log linear analysis was conducted on participants' decision to confess. As seen in Table 2, this analysis revealed a strong and predictable main effect of guilt, G^2 (1, $N = 59$) = 26.33, $p < .001$, 95% CI .10 to .50, as the confession rate increased substantially from 26.7% overall in the innocent condition (8/30) to 89.7% in the guilty condition (26/29). There was also a significant main effect of the bluff tactic, G^2 (1, $N = 59$) = 3.87, $p < .05$, 95% CI .00 to .15, such that the confession increased from 44.8% in the control condition (13/29) to 70.0% when the bluff was used (21/30). Importantly, these main effects were qualified by a significant two-way interaction, G^2 (4, $N = 59$) = 39.40, $p < .001$, 95% CI .16 to .64. A ceiling effect for confessions was obtained in the guilty condition, which limited the potential to obtain an effect for the bluff on guilty participants, G^2 (1, 29) = .31, 95% CI .00 to .62. In contrast, and replicating the effect obtained in the first two studies, the confession rate increased substantially in the innocent condition from 0 to 50% with the addition of the bluff, G^2 (1, 30) = 12.61, $p < .001$, 95% CI .03 to .30.

Table 2 Confession rates and diagnosticity ratios as a function of guilt and bluff conditions in Study 3

Tactics	Suspect condition		
	Guilty (%)	Innocent (%)	Diagnosticity
Control	87^{ab}	00^{c}	87.00*
Bluff	93^{b}	50^{a}	1.86

Note: Percentages not sharing a common superscript differ at $p < .05$

* A constant of 1 was added to the denominator to allow for the calculation of diagnosticity

Given the importance of identifying interrogation techniques that yield a high rate of true confessions and a low rate of false confessions, we calculated diagnosticity by the ratio of true and false confessions in the bluff and control conditions. As a consequence of its dual effects on guilty and innocent participants, the result confirmed that the bluff tactic severely diminished the diagnosticity of outcomes. In the control condition, a high diagnosticity ratio was achieved, with confessions from 87% of guilty participants and none of the innocents. In the bluff condition, however, diagnosticity dropped to 1.86 as a result of confessions from 93% of guilty participants and 50% of the innocents. Clearly, the bluff tactic used in this study, and the false confessions it induced, made it more difficult to differentiate between guilty and innocent participants (see Table 2).

As in Study 2, participants' self-reported explanations for confessions and denials were analyzed. Two participants had to be excluded from analyses because of audiotape failures, so a total of 57 interviews were coded by two independent raters ($\kappa = .852$). Disagreements were resolved by discussion. The results indicated even though the potential for a bluff effect could not be assessed in the guilty condition because of the ceiling effect, 45% of guilty confessors in the bluff condition cited the bluff as a reason for their decision to confess (e.g., "Mm, it was easier to sign it because then there's no point of me really lying to you, then you see the camera, I mean I knew I did it anyway."). Turning to the innocent condition, we found that 88% of innocent confessors confronted with the bluff cited this tactic as a reason for their decision to confess, as we had predicted—but so did 75% who refused to sign the statement. Consistent with Study 2, these self-reports both

support the interrogation theory that innocent suspects are protected by the bluff tactic and the contradictory hypothesis that they are adversely affected by it (e.g., "it made it easier [to sign] because I had nothing to hide. The cameras would prove it.").

General Discussion

Guided by recent theorizing on the phenomenology of innocence, the present research was designed to investigate the impact of the bluff tactic on confession rates in the laboratory. Using the computer crash paradigm, Study 1 showed that the bluff substantially increased the false confession rate, by 60%, as much as did the presentation of false evidence. Study 2 replicated this effect, again using the computer crash paradigm, and added self-report evidence indicating that most participants in the bluff condition consciously cited that tactic, and the resulting inference of future exoneration, as a reason for their false confessions. Using the cheating paradigm, Study 3 replicated the bluff effect on innocent participants even for a willful act of cheating that may have violated the university honor code. In short, the bluff effect on innocent laboratory participants is unequivocal. Indeed, as questions concerning generalizability can be raised about laboratory studies (e.g., where students participate in a lesser stakes task), it is noteworthy that the bluff effect hypothesis was born not from theory but from the retrospective self-reports of real false confessors.

Although the bluff is as potent at producing false confessions as the presentation of false evidence obtained in our first experiment and elsewhere, it does so by a more circumscribed mechanism. Research on the false evidence effect has shown that a subsample of innocent confessors become convinced of their guilt and internalize responsibility for an outcome they did not produce (e.g., Horselenberg et al., 2003; Kassin & Kiechel, 1996; Redlich & Goodman, 2003). In the present studies, however, all innocent participants—fully aware of their lack of culpability—used the alleged evidence contained in the bluff to extricate themselves from the situation in an act of compliance. Rather than wait for the computer keystrokes or videotaped evidence to be retrieved, these participants decided to sign an admission of guilt, fully expecting to be exonerated later (interestingly, one participant refused to sign in the bluff condition because she believed that the bluff was a lie that the experimenter had told to convince her to confess).

The willingness of innocent participants to confess was evident not only in the computer crash paradigm used in Experiments 1 and 2 but also in the cheating paradigm used in Experiment 3. The latter results are particularly disconcerting. One virtue of this paradigm is that participants are pressured to confess to an act that is not only prohibited but also willful and of consequence. Indeed, the zero confession rate in the innocent control condition of this experiment suggests that the situation strongly inhibited confessions from innocent participants, many of whom preferred to defend themselves face-to-face with the professor whose research they were accused of compromising. That half of all innocents in the bluff condition chose to confess indicates that this tactic served not as a threat of inculpation but rather as a promise of future exoneration—which, paradoxically, made it easier to confess.

This paradoxical effect of the bluff on innocent participants contributes to recent work indicating that innocence can put those accused at risk for the decisions they make in an interrogation setting. Research has shown that people who stand falsely accused waive their *Miranda* rights to silence and to counsel (Kassin & Norwick, 2004) and to a lineup (Holland et al., 2005). Then once questioned by police, they are fully cooperative and forthcoming (Hartwig et al., 2005, 2006) and exhibit less physiological arousal than mock criminals on critical items of a concealed information test even when informed of those details (Zvi et al., 2010). Kassin (2005) suggested that innocent people harbor a naïve "phenomenology of innocence" that is rooted in a generalized belief in a just world in which human beings "get what they deserve" and "deserve what they get" (Lerner, 1980) and in an "illusion of transparency" whereby people overestimate the extent to which their true inner states are discernible to others (Gilovich, Savitsky, & Medvec, 1998; Miller & McFarland, 1987). By confirming previous anecdotal reports that the bluff technique—a lawful and apparently benign interrogation tactic that should trap only perpetrators into confession—can also induce false confessions, the present experiments extend this concept of innocence as a risk factor an important next step.

These findings are particularly problematic in light of other research showing that police and lay jurors cannot easily distinguish between true and false confessions (Kassin, Meissner, & Norwick, 2005); that many actual false confessions contain accurate crime facts that were not in the public domain (Garrett, 2010); that many confessions contain vivid sensory descriptions of what, how, and why the suspect committed the crime, as well as expressions of voluntariness, apologies, and remorse (Appleby, Hasel, Shlosberg, & Kassin, 2009); and that jurors and judges do not fully discount confessions, even when they are elicited through coercive methods (Kassin & Sukel, 1997; Wallace & Kassin, 2009). Although this hypothesis is yet to be tested, we would predict that bluff-induced false confessions are particularly likely to be misperceived to be voluntary and as true because of the apparently benign nature of the deception involved compared to the outright lies that comprise the false evidence ploy. In short, false

confessions induced by the bluff tactic are unlikely to be detected by police, prosecutors, judges, juries, and others.

Taken together, the present studies convincingly demonstrate that use of the bluff tactic in an interrogation can induce compliant false confessions from innocent people. Importantly, however, additional research is needed to reassess the predicted effectiveness of the bluff on the true confession rates of perpetrators and, hence, the diagnosticity of outcomes. We sought to test this hypothesis in Experiment 3 but a ceiling effect curtailed our ability to do so, with 87% of participants confessing in the guilty control condition—compared to 93% in the guilty bluff condition. Although we cannot be certain of the reason for this high baseline rate of true confessions, several participants in the guilty condition said afterward that they confessed out of fear that the confederate, from whom they were separated for questioning, would implicate them. With the confederate thus serving as something of an "accomplice" in the cheating paradigm, it appears that our guilty participants—like many perpetrators—felt trapped by the potential other form of incriminating evidence (other than the bluffed recording). In fact, many participants may have felt pressured by a sense that they were in a prisoner's dilemma situation. Although we presented no incentive for being the first to confess, some participants said they wanted to ensure that we knew they had only aided the confederate upon request and did not initiate the cheating.

At this point, it seems clear that the bluff technique does not insulate innocent suspects from the risk of false confession—either in absolute terms or relative to the more controversial false evidence ploy. To the contrary, in light of the phenomenology of innocence, which can lead people who stand falsely accused to confess out of an expectation of future exoneration, the bluff itself puts innocent people at risk and should be approached with great caution. In this vein, further research is still needed—using laboratory paradigms in which guilt–innocence and interrogation tactics are independently varied (see Meissner et al., 2010)—to help build interrogation models that both minimize false confessions and produce more diagnostic outcomes.

Acknowledgments The research reported in this manuscript was supported by funds from the City University of New York to the second author. We are indebted to Jennifer Schell for her invaluable role as experimenter in Study 3. We also want to thank Amanda Baird, Rebecca Dryden-Shepherd, and Vanessa Materko, for their role as confederates.

References

Appleby, S., Hasel, L. E., & Shlosberg, A., & Kassin, S. M. (2009). False confessions: A content analysis. Paper presented at the American Psychology-Law Society, San Antonio, TX.

Firstman, R., & Salpeter, J. (2008). *A criminal injustice. A true crime, a false confession, and the fight to the free Marty Tankleff*. New York, NY: Ballantine Books.

Frazier v. Cupp, 394 U.S. 731 (1969).

Garrett, B. L. (2008). Judging innocence. *Columbia Law Review, 108*, 55–142.

Garrett, B. L. (2010). The substance of false confessions. *Stanford Law Review, 62*, 1051–1119.

Gilovich, T., Savitsky, K., & Medvec, V. H. (1998). The illusion of transparency: Biased assessments of others' ability to read one's emotional states. *Journal of Personality and Social Psychology, 75*, 332–346. doi:10.1037/0022-3514.75.2.332.

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

Gudjonsson, G. H., & Sigurdsson, J. F. (1999). The Gudjonsson Confession Questionnaire-Revised (GCQ-R): Factor structure and its relationship with personality. *Personality and Individual Differences, 27*, 953–968. doi:10.1016/S0191-8869(98)00278-5.

Hartwig, M., Granhag, P. A., Strömwall, L. A., & Kronkvist, O. (2006). Strategic use of evidence during police interviews: When training to detect deception works. *Law and Human Behavior, 30*, 603–619. doi:10.1007/s10979-006-9053-9.

Hartwig, M., Granhag, P. A., Strömwall, L. A., & Vrij, A. (2005). Detecting deception via strategic disclosure of evidence. *Law and Human Behavior, 29*, 469–484. doi:10.1007/s10979-005-5521-x.

Holland, L., Kassin, S. M., & Wells, G. L. (2005, March). Why suspects waive the right to a lineup: A study in the risk of actual innocence. Poster presented at the Meeting of the American Psychology-Law Society, San Diego, CA.

Horselenberg, R., Merckelbach, H., & Josephs, S. (2003). Individual differences, false confessions: A conceptual replication of Kassin, Kiechel (1996). *Psychology, Crime, and Law, 9*, 1–18. doi:10.1080/10683160308141.

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. (2005). On the psychology of confessions: Does innocence put innocents at risk? *American Psychologist, 60*, 215–328. doi:10.1037/0003-066X.60.3.215.

Kassin, S. M. (2007). Internalized false confessions. In M. P. Toglia, J. D. Read, D. F. Ross, & R. C. L. Lindsay (Eds.), *The handbook of eyewitness psychology, Vol. 1: Memory for events* (pp. 175–192). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Kassin, S. M., Drizin, S. A., Grisso, T., Gudjonsson, G. H., Leo, R. A., & Redlich, A. D. (2010). Police-induced confessions: Risk factors and recommendations. *Law and Human Behavior, 34*, 3–38. doi:10.1007/s10979-009-9188-6.

Kassin, S. M., Goldstein, C. J., & Savitsky, K. (2003). Behavioral confirmation in the interrogation room: On the dangers of presuming guilt. *Law and Human Behavior, 27*, 187–203. doi:10.1023/A:1022599230598.

Kassin, S. M., & Gudjonsson, G. H. (2004). The psychology of confession evidence: A review of the literature and issues. *Psychological Science in the Public Interest, 5*, 35–69. doi:10.1111/j.1529-1006.2004.00016.x.

Kassin, S. M., & Kiechel, K. L. (1996). The social psychology of false confessions: Compliance, internalization, and confabulation. *Psychological Science, 7*, 125–128. doi:10.1111/j.1467-9280.1996.tb00344.x.

Kassin, S. M., Leo, R. A., Meissner, C. A., Richman, K. D., Colwell, L. H., Leach, A.-M., et al. (2007). Police interviewing and interrogation: A self-report survey of police practices and beliefs. *Law and Human Behavior, 31*, 381–400. doi:10.1007/s10979-006-9073-5.

Kassin, S. M., Meissner, C. A., & Norwick, R. J. (2005). 'I'd know a false confession if I saw one': A comparitive study of college students and police investigators. *Law and Human Behavior, 29*, 211–227. doi:10.1007/s10979-005-2416-9.

Kassin, S. M., & Norwick, R. J. (2004). Why suspects waive their Miranda rights: The power of innocence. *Law and Human Behavior, 28*, 211–221. doi:10.1023/B:LAHU.0000022323.74584.f5.

Kassin, S. M., & Sukel, H. (1997). Coerced confession and the jury: An experimental test of the "Harmless Error" rule. *Law and Human Behavior, 21*, 27–46. doi:10.1023/A:1024814009769.

Kassin, S. M., & Wrightsman, L. S. (1985). Confession evidence. In S. M. Kassin & L. S. Wrightsman (Eds.), *The psychology of evidence and trial procedures* (pp. 67–94). London: Sage.

Leo, R. A. (1996). Inside the interrogation room. *The Journal of Criminal Law and Criminology, 86*, 266–303. doi:10.2307/1144028.

Leo, R. A. (2008). *Police interrogation and American justice*. Cambridge: Harvard University Press.

Lerner, M. J. (1980). *The belief in a just world*. New York: Plenum.

Loftus, E. F. (2005). Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning & Memory, 12*, 361–366. doi:10.1101/lm.94705.

Meissner, C. A., Russano, M. B., & Narchet, F. M. (2010). The importance of a laboratory science for improving the diagnostic value of confession evidence. In G. D. Lassiter & C. A. Meissner (Eds.), *Police interrogations and false confessions: Current research, practice, and policy recommendations* (pp. 111–126). Washington, DC: American Psychological Association. doi:10.1037/12085-007.

Miller, D. T., & McFarland, C. (1987). Pluralistic ignorance: When similarity is interpreted as dissimilarity. *Journal of Personality and Social Psychology, 53*, 298–305. doi:10.1037/0022-3514.53.2.298.

Missouri v. Johnson, No. 98–5364 (2001).

Moston, S., Stephenson, G. M., & Williamson, T. M. (1992). The effects of case characteristics on suspect behaviour during police questioning. *British Journal of Criminology, 32*, 23–40.

Nash, R. A., & Wade, K. A. (2009). Innocent but proven guilty: Using false video evidence to elicit false confessions and create false beliefs. *Applied Cognitive Psychology, 23*, 624–637. doi:10.1002/acp.1500.

Redlich, A. D., & Goodman, G. S. (2003). Taking responsibility for an act not committed: Influence of age and suggestibility. *Law and Human Behavior, 27*, 141–156. doi:10.1023/A:1022543012851.

Russano, M. B., Meissner, C. A., Narchet, F. M., & Kassin, S. M. (2005). Investigating true and false confession within a novel experimental paradigm. *Psychological Science, 16*, 481–486.

Santos, F. (2006, September 20). DNA evidence frees a man imprisoned for half his life. *The New York Times*, A1.

Swanner, J. K., Beike, D. R., & Cole, A. T. (2010). Snitching, lies and computer crashes: An experimental investigation of secondary confessions. *Law and Human Behavior, 34*, 553–565. doi:10.1007/s10979-008-9173-5.

Wallace, B. D., & Kassin, S. M. (2009). Harmless error analysis: Judges' performance with confession errors. Paper presented at the American Psychology-Law Society, San Antonio, TX.

Wilson, T. D., & Nisbett, R. E. (1978). The accuracy of verbal reports about the effects of stimuli on evaluations and behavior. *Social Psychology, 41*, 118–131. doi:10.2307/3033572.

Zulawski, D. E., & Wicklander, D. E. (2002). *Practical aspects of interview and interrogation* (2nd ed.). Boca Raton, FL: CRC Press.

Zvi, L., Nachson, I., & Elaad, E. (2010). Effects of coping and cooperative behavior of guilty and informed innocent participants on the detection of concealed information. Poster presented at the Meeting of the European Association of Psychology and Law, Goteborg, Sweden.