Law and Human Behavior

Police Reports of Mock Suspect Interrogations: A Test of Accuracy and Perception

Saul M. Kassin, Jeff Kukucka, Victoria Z. Lawson, and John DeCarlo

CITATION

Kassin, S. M., Kukucka, J., Lawson, V. Z., & DeCarlo, J. (2016, December 12). Police Reports of Mock Suspect Interrogations: A Test of Accuracy and Perception. *Law and Human Behavior*. Advance online publication. http://dx.doi.org/10.1037/lhb0000225

Police Reports of Mock Suspect Interrogations: A Test of Accuracy and Perception

Saul M. Kassin
John Jay College of Criminal Justice

Jeff Kukucka
Towson University

Victoria Z. Lawson
Institute for State and Local Governance of the City University of New York

John DeCarlo
University of New Haven

A 2-phased experiment assessed the accuracy and completeness of police reports on mock interrogations and their effects on people's perceptions. In Phase 1, 16 experienced officers investigated a mock crime scene, interrogated 2 innocent suspects—1 described by the experimenter as more suspicious than the other—and filed an incident report. All 32 sessions were covertly recorded; the recordings were later used to assess the reports. In Phase 2, 96 lay participants were presented with a brief summary of the case and then either read 1 police report, read 1 verbatim interrogation transcript, or listened to an audiotape of a session. Results showed that (a) Police and suspects diverged in their perceptions of the interrogations; (b) Police committed frequent errors of omission in their reports, understating their use of confrontation, maximization, leniency, and false evidence; and (c) Phase 2 participants who read a police report, compared to those who read a verbatim transcript, perceived the process as less pressure-filled and were more likely to misjudge suspects as guilty. These findings are limited by the brevity and low-stakes nature of the task and by the fact that no significant effects were obtained for our suspicion manipulation, suggesting a need for more research. Limitations notwithstanding, this study adds to a growing empirical literature indicating the need for a requirement that all suspect interrogations be electronically recorded. To provide a more objective and accurate account of what transpired, this study also suggests the benefit of producing verbatim transcripts.

Keywords: interrogations, suspects, police reports, accuracy

Supplemental materials: http://dx.doi.org/10.1037/lhb0000225.supp

Over the years, research on police interrogations, confessions, and their role in known cases of wrongful conviction has animated calls for reform (for reviews, see Gudjonsson, 2003; Kassin, 1997, 2005, 2012; Kassin & Gudjonsson, 2004; Lassiter & Meissner, 2010; for an official white paper, see Kassin et al., 2010). Many such reform efforts have been aimed at protecting highly vulnerable populations (e.g., juveniles, people with intellectual or mental health impairments) and at curtailing the use of coercive interrogation practices (e.g., presentations of false evidence, minimization tactics that imply leniency). Perhaps the most significant proposed safeguard is to require the electronic recording of interrogations—the entire process, not just the confession. As stated in the AP-LS white paper: "Without equivocation, our most essential recommendation is to lift the veil of secrecy from the interrogation process in favor of the principle of transparency" (Kassin et al., 2010, p. 25).

There is a perennial debate concerning the recording of suspect interviews and interrogations (for an overview, see Drizin & Reich, 2004). As an historical matter, the practice has drawn strong resistance from many federal, state, and local police professionals (e.g., Boetig, Vinson, & Weidel, 2006)—especially those trained by John Reid & Associates, which, until recently, had steadfastly opposed the recording of interrogations (Inbau, Reid, Buckley, & Jayne, 2001). The bases of opposition have varied. Some have opposed recording on pragmatic and logistical grounds—citing the scope of such a requirement; financial costs; the evidentiary consequences of a failure to comply, for example, due to equipment malfunction; and issues of consent, especially in two-party consent states. Others have expressed concern over how recording might alter the behavior of both police and suspects during interrogation and the subsequent decision-making of judges and juries

(for an overview of these arguments, see Sullivan, 2008; Thurlow, 2005).

At the same time, a policy of mandatory recording has received support from various organizations (e.g., American Bar Association, 2004; American Psychological Association, 2014; Buckley & Jayne, 2005; The Justice Project, 2007), as well as from surveys of police investigators across the United States and Canada (Geller, 1993; Kassin et al., 2007). Sullivan (2004) interviewed police from hundreds of departments that recorded custodial interrogations and consistently found that they fully embraced the practice. Respondents cited numerous benefits of recording, namely that it allowed detectives to focus on the suspect rather than on concurrent note-taking; that it allowed them to later review the suspect's answers to questions for any incriminating comments that had initially gone unnoticed; that it lessened the need for detectives to defend their interrogation practices in court; and that it enhanced public trust in law enforcement (also see Sullivan, Vail, & Anderson, 2008). Importantly, the U.S. Department of Justice recently reversed its longstanding opposition to recording by establishing the presumptive requirement that federal law enforcement agencies (including the FBI) videotape the custodial interrogations of felony suspects (Schmidt, 2014).

That police who record interrogations report high levels of satisfaction is a vital data point in efforts to reform majority practice. But what are the actual effects? There are two primary sets of reasons for the proposed recording requirement. The first is the expectation that the practice of recording will increase accountability and discourage the use of coercive interrogation tactics, thus reducing the risk to all suspects. To test this hypothesis, Kassin, Kukucka, Lawson, and DeCarlo (2014) conducted an experiment in a police station, in which 61 experienced investigators interrogated a male suspect who was either guilty or innocent of a mock theft. Before each interrogation, the investigator either was or was not informed that their session would be surreptitiously recorded. These recordings were later coded for the use of various high-pressure tactics designed to elicit a confession. As predicted, camera-informed interrogators were less likely than their uninformed counterparts to use both maximization and minimization tactics; they were also judged by suspects—who were not told of the camera's presence—as trying less hard to obtain a confession. These findings suggest that recording can affect the process of interrogation—namely, by inhibiting the use of coercive tactics.

A second purported benefit of recording interrogations is to provide an accurate factual record of the interrogation behavior of police and suspects. Perhaps the most frequently invoked argument is that it is the most effective way to memorialize the process by which a statement was taken and, hence, increase the fact-finding accuracy of prosecutors (who decide whether to charge a suspect), judges (who rule on whether a confession was voluntary or coerced), and juries (who determine whether a confession is credible and hence whether the confessor is guilty or innocent). In current practice, whereby detectives take contemporaneous or retrospective notes, disputes often arise as to whether *Miranda* rights were administered and waived in a timely manner; whether the suspect was cooperative or evasive; whether police made or implied promises or threats, or lied about evidence; and, importantly, whether the details contained within a confession originated from the suspect. Disputes over this latter issue can prove devastating. In a descriptive analysis of 38 false confessions from the Innocence Project, Garrett

(2010) found that 36 were "contaminated," containing accurate crime details allegedly known only to the perpetrator—details not in the public domain, but known to police, that the innocent confessor could not have produced without exposure to secondhand information. The result: An increased likelihood of conviction, as the presence of details in a confession enhances perceptions of its credibility (Appleby, Hasel, & Kassin, 2013).

In lieu of electronic recordings, the Federal Rules of Evidence (2015) provide that a police witness may use a personally prepared report concerning an interview or interrogation, including what the defendant said, to refresh his or her recollection while testifying (FRE 612). Yet the accuracy of these reports, which is often in dispute, has never been tested in the context of a suspect interrogation. The question we posed in the present research is: What does recording reveal about the memorial accuracy of police reports of interrogations, and with what effect on prospective fact finders?

Basic research on memory for conversation content suggests that this process may be fraught with bias and error. Neisser (1981) highlighted this problem in a case study in which he analyzed John Dean's high-stakes memory of Watergate-related conversations with President Nixon. Dean testified as to his recollections with specificity and confidence. Yet when tapes of oval office conversations were discovered, Neisser discovered that although Dean was generally correct about what happened (e.g., that there was a cover-up), his memory of specifics was often distorted (e.g., overestimating his own role). This case stands in contrast to research showing that active involvement in a conversation tends to enhance memory for content; for example, people recall what they said better than what they read or heard (MacLeod, Gopie, Hourihan, Neary, & Ozubko, 2010).

Apart from the potential for intrusion and omission errors in memory for conversations, research shows that memories of *content* and *context* are stored independently, thereby increasing the risk of source monitoring confusion (Johnson, Hashtroudi, & Lindsay, 1993). Moreover, such errors often reflect the operation of cognitive confirmation biases (Kleider, Pezdek, Goldinger, & Kirk, 2008; Schacter, 2001). In a study that demonstrated the problem in a forensic context, Lamb, Orbach, Sternberg, Hershkowitz, and Horowitz (2000) took a sample of 20 interviews of alleged child sex abuse victims and compared interviewers' contemporaneous notes against audiotapes of these sessions. The interviewers' notes proved inadequate, as they failed to mention 57% of their own utterances and 25% of details that the children provided. Moreover, the notes often contained serious source attribution errors, such as citing the children rather than their own prompting questions as the source of details that were disclosed. Lamb et al. (2000) concluded: "Even when they made contemporaneous verbatim notes, these investigators tended to understate their role in eliciting the information" (p. 705; see also Bruck, Ceci, & Francoeur, 1999).

The consequences of source attribution errors were realized by former D.C. Detective James Trainum (2007) who—in an article entitled "I took a false confession so don't tell me it doesn't happen!"—described a case in which a former suspect who had confessed was later exonerated:

> Years later, during a review of the videotapes, we discovered our
> mistake. We had fallen into a classic trap. We believed so much in our

suspect's guilt that we ignored all evidence to the contrary. To demonstrate the strength of our case, we showed the suspect our evidence, and unintentionally fed her details that she was able to parrot back to us at a later time. It was a classic false confession case and without the video we would never have known.

To further complicate matters of recollection, there is reason to believe that suspects likewise cannot be trusted to provide accurate accounts. Unlike most social interactions, police interrogations are exceptionally stressful events for the accused, and this level of stress can produce deleterious effects on memory retrieval (Deffenbacher, Bornstein, Penrod, & McGorty, 2004). In a study that illustrates this point, Morgan et al. (2004) randomly assigned trainees undergoing military survival training to endure a realistic high-stress or low-stress mock interrogation. One day later, many of those in the high-stress condition could not even identify their interrogator from a lineup.

In light of the research literature on memory for conversations, the acceptance in court of note taking in lieu of electronic recording, and the significance of the issue for policy and practice, the present research was designed to assess the memorial accuracy of police reports of suspect interrogations. In an experiment conducted at a large Northeastern police station, a sample of experienced investigators examined a mock crime scene, interrogated two innocent male suspects, and submitted a report on their interrogations. For each investigator, expectations were varied from one session to the next, with one of the two suspects presented via demeanor cues as having behaved suspiciously. Unbeknownst to these investigators, all sessions were covertly audio recorded. By later comparing police reports of the sessions with the actual tapes, we addressed two questions. First, what does audio recording reveal about substantive accuracy of police accounts, as measured by intrusion errors, omission errors, and source monitoring confusion, and what does it reveal about the extent to which suspicion elicits confirmation bias in police reports of interrogations? Second, with regard to the fact finder, how might access to the unabridged verbatim transcripts influence others' perceptions of the interrogations and suspects compared to reading only the police reports?

Phase 1 Method

Participants

Phase 1 participants were 18 investigators of various ranks from a large Northeastern police department and 36 male community members recruited via Craigslist to serve as mock suspects. Police participants were recruited during daily roll calls and later via word-of-mouth once the study was ongoing. As noted in the consent form, the stated purpose of the research was "to study the way police investigators conduct suspect interviews and interrogations, and write informative incident reports on their work." Data from one police participant were excluded after he expressed suspicion that his session had been recorded; a second police participant failed to submit the required written report following his session. Hence, the final sample consisted of $N = 16$ police participants (eight officers, two detectives, and six sergeants) who conducted and reported on a total of 32 suspect interrogations. All sessions were conducted in a vacant office at the police station.

Police participants, half of whom were female, ranged in age from 26 to 55 ($M = 43.94$, $SD = 7.46$) and had an average of 16.33 years of law enforcement experience ($SD = 6.00$). Half had received formal training in suspect interviewing and interrogation; 62.50% estimated that they had conducted over 100 suspect interviews during their careers (if a range was provided, we converted the estimate to the midpoint of that range; the overall median was 250). All suspects were male and ranged in age from 18 to 62 ($M = 34.91$, $SD = 13.86$). A total of 46.88% had previously been arrested; 25% had been suspect-interviewed by police; 28.13% had been convicted of a crime.

Design

Within each session, one police participant investigated a staged crime scene involving a theft and then interrogated two male suspects, both of whom were factually innocent. After looking at the crime scene but prior to interrogating the first suspect, they were told that the first or second of the two suspects (in counterbalanced order) had acted suspiciously upon learning that he would be questioned about the theft. Phase 1 thus employed a two-group (Suspicion: Present vs. Absent) within-subjects design.

Procedure

Two experimenters were involved in each session—one to meet and instruct the police participant, and the other to separately meet and instruct the two suspects. Each session began with a preexperiment questionnaire, at which time a theft was said to have occurred. This was followed by a police crime scene investigation; two sequential suspect interrogations; and postinterrogation questionnaires. After each session, the police participant was asked to produce and submit a written report within 48 hours that detailed his or her crime scene investigation and interrogations.

Preexperiment questionnaires and mock theft. Upon arrival, the police participant was escorted to an interview room; the two suspects were escorted to a separate waiting room. After giving informed consent, the police participant indicated his or her age and gender and answered several questions concerning background and training.

After suspect participants gave their informed consent, an experimenter directed but did not accompany them, one-at-a-time, to another room where they were to complete a preexperiment questionnaire. The questionnaire asked suspects to indicate their age and whether they had ever been interviewed by police, arrested, or convicted of a crime. On their way to and from this room, suspects passed an unattended briefcase in the hallway that had been staged to look as though a theft had occurred. The briefcase zipper was left ajar, and on the floor next to it were an open lock (which had ostensibly been removed from the briefcase) and a wallet with no cash inside. Suspects were instructed to take notice of—but not to touch—the briefcase. All suspects, therefore, had exposure to the scene of the mock theft but were factually innocent.

Once both suspects had completed the questionnaire and returned to the waiting room, the experimenter informed them that they were the targets of an investigation into a theft committed at the police station and that they would soon be

interviewed by a detective. Suspects were paid $20 for their participation prior to being interrogated and told that they would receive a $15 bonus if they convinced the detective of their innocence. If they did not succeed, they were told that they would have to return for a second session to receive the additional payment (in actuality, all suspects were paid the full $35 after the session).

Crime scene investigation.   At that point, the second experimenter informed the police participant that an unknown sum of cash had been stolen from a briefcase in the hallway and that the crime scene had been left exactly as it was found. In the context of a lengthy instruction, police were asked to investigate the crime scene, question two suspects who were apprehended nearby around the time that the theft occurred, and solve the crime. The experimenter explained that both suspects had walked past the briefcase to complete paperwork in a nearby room and were out of view when they did so. We provided no inculpatory evidence against the suspects aside from the fact that the money was missing and both were known to have been alone in the vicinity at the time. Police were assured that no actual crime had taken place and that interrogation must be terminated if a participant suspect wanted to stop (the verbatim instructions are available as online supplemental material).

The police participant was then taken to the crime scene, provided with a notepad and digital camera, and given 5 minutes to investigate the area. Afterward, the experimenter escorted the police participant back to the interrogation room and inquired as to whether he or she wanted hard copies of their crime scene photographs for use during the interrogation. If they did, these images were printed, in color, on 4 × 6 in. photo paper.

Suspicion manipulation.   While escorting the police participant back to the interrogation room, the experimenter reiterated that he or she would question two theft suspects. In counterbalanced order, the experimenter added that:

> The first [or second] guy you're going to question seemed okay, but the second [or first] guy was acting really strange. He looked nervous: He kept pacing back and forth and wouldn't make eye contact with me while I was talking to him.

Given that a mock crime was said to have been committed, this comment about demeanor was designed to selectively raise each police participant's relative a priori suspicion toward one of the two suspects.

Interrogations and postinterrogation questionnaires.   Police were told that they would have 20 to 30 min to interrogate each suspect, after which they were to return the suspect to the waiting room. If an interrogation was still ongoing at 20 min, the experimenter knocked on the door to signal to the police participant that it was time to wrap up the interrogation. Suspects were randomly assigned to be interrogated either first or second. All interrogations were surreptitiously audio recorded through a digital voice recorder hidden among office supplies on a desk. Neither police participants nor suspects were preinformed of the fact that the interrogations would be recorded.

After each interrogation, the suspect returned to the waiting room, and both the police participant and suspect completed a self-report questionnaire concerning their perceptions of the interrogation experience. Police participants completed this questionnaire twice, once after each of their interrogations. After the suspect completed his questionnaire, he was fully debriefed, paid, and dismissed.

Police reports.   After completing the second postinterrogation questionnaire, police were instructed on how to prepare and submit Incident Reports of their investigation, which were due within 48 hours. To ensure that investigators took the task seriously, these instructions stipulated that reports should be approximately three to five typed pages in length, single spaced, and should consist of three sections: a summary of their crime scene analysis (one page), an account of the first suspect interrogation (1–2 pages), and an account of the second suspect interrogation (1–2 pages). The instructions also specified the sorts of details that should be included—including key questions asked during each interrogation, suspects' answers to these questions, descriptions of each suspect's demeanor, and any indications or impressions of each suspect's involvement. Upon receipt of these reports, police were debriefed and paid $100 for their participation.

Dependent Measures

Postinterrogation questionnaires.   Immediately after each interrogation, police and suspect participants completed a self-report questionnaire consisting of 12 parallel items that measured their perceptions of the interrogation. First, police indicated whether they believed the suspect to be guilty or innocent and indicated their confidence in this impression on a scale ranging from 1 (*not at all*) to 10 (*very*). These two items were later combined to form a guilt-confidence composite score that could range from −10 (*highly confident guilty judgment*) to +10 (*highly confident innocent judgment*). Police also gave continuous ratings of how credible the suspect's denials were, how knowledgeable the suspect seemed to be about the crime, how cooperative the suspect was, how hard they tried to get the suspect to confess, how friendly they were toward the suspect, how confrontational they were toward the suspect, how anxious the suspect was, and how stressful they thought the interrogation was for the suspect. All ratings were made on a scale that ranged from 1 (*not at all*) to 10 (*very*). Finally, police gave two dichotomous yes/no judgments as to whether the suspect had made any suspicious remarks and/or any admissions of guilt during the interrogation.

Suspects simultaneously answered parallel questions. First, they indicated whether they believed the interrogator would perceive them to be guilty or innocent and rated their confidence in that judgment on a 10-point scale. Once again, these two items were combined to form a composite score that could range from −10 (*highly confident that they would be judged guilty*) to +10 (*highly confident that they would be judged innocent*). Suspects also rated how credible their denials were, how much knowledge they had about the crime, how cooperative they were, how hard the interrogator tried to get them to confess, how friendly the interrogator was, how confrontational the interrogator was, how anxious they were, and how stressful the interrogation was for them. These continuous ratings were made on a scale ranging from 1 (*not at all*) to 10 (*very*). Suspects also gave dichotomous yes/no judgments as to whether they had made any suspicious remarks and/or any admissions of guilt. Two suspects (6.25%) neglected to answer one or more of these items, and thus their data are missing from the relevant analyses.

Interrogation tactics. Audio recordings of all 32 interrogations were transcribed by a professional transcription service. The transcripts were then coded by two independent coders who were blind to our suspicion manipulation and had prior experience coding interrogation transcripts. On the basis of a previously published factor analysis of self-reported tactic use in a survey of police (Kassin et al., 2007), which formed the basis of a coding scheme introduced by Kassin et al. (2014), we coded transcripts for the presence of 16 interrogation tactics organized into five categories (see Table 1).

For three of the five categories (confrontation, maximization, and leniency), coders counted the number of times each individual tactic was used. For some analyses, these counts were later coded as absent (i.e., a count of zero) or present (i.e., any nonzero count). We also computed the sum of tactic uses within each category, enabling us to create a dichotomous measure of whether or not police used at least one tactic from a given category. For the other two categories (false evidence and miscellaneous), coders merely noted whether each tactic was present or absent.

Confrontation tactics involved calling the suspect a liar (e.g., "You just lied to me," "You're pretty much lying to an authority figure you know"), pointing out inconsistencies in the suspect's story (e.g., "You're flipping your story," "How do you know there was a card in the wallet? You said you just looked at it"), outright accusing the suspect of committing the crime (e.g., "You did it," "Okay, well I think that you took the money"), expressing disbelief in the suspect's story (e.g., "That's hard to believe," "I can tell you right now just from the evidence that I have that your story is not believable"), and interrupting the suspect's attempts to maintain his innocence (as denoted by midsentence hyphens in the suspect portions of the transcripts).

Maximization tactics include threatening the suspect with negative consequences for not confessing (e.g., "If it comes back to you, you're done," "If you don't tell me up front and then I have to determine that by evidentiary matter then I slam you in court, okay?"), and exaggerating the seriousness of the offense (e.g., "The amount of cash in the wallet can bring the charges from now a theft to a much higher degree," "There's going to be an additional charge for hindering prosecution").

Leniency tactics include developing minimization themes that excuse, justify, or otherwise downplay the crime, statements that may lead people to infer leniency in punishment (e.g., "This is something, it's so minor," "It doesn't make you a bad person, bro") and explicit offers of leniency or immunity in exchange for a confession (e.g., "I got a little bit of juice over at the courthouse and if you did do it, if you're honest and up front with me, I can work with you," "I'll get you probation on this," "I am offering you a deal here then, you know").

False evidence tactics include the bluff (the assertion that there is evidence to be harvested without the added claim that it implicates the suspect; e.g., "Understand I am going to check the cameras," "That little padlock and key that you saw, that has been submitted for DNA analysis") and specific false claims about the existence of actual evidence (e.g., "I already told you there is one person that ID'd you right?" "We have video of you going in that wallet").

In addition to these basic Reid-technique approaches, a range of miscellaneous tactics were coded as well—including irrelevant small talk aimed at establishing a rapport (e.g., "What kind of food you like cooking?" "What kind of work do you do?"), asking the suspect to implicate someone else (e.g., "Did you see any of the other guys like [John], his brother, whoever, anybody with money take money out of their pockets, anything?" "What about the, um, older gentleman, was he referencing anything about money?"), encouraging the suspect to admit to other illegal behaviors (e.g., "Are you on probation or anything?" "Okay have you ever been

Table 1

Presence and Frequency of 16 Categorized Interrogation Tactics Across 32 Interrogations

Interrogation tactics	% Sessions used	M (SD) of times used
Confrontation	**56.25**	**1.63 (1.98)**
Calling the suspect a "liar" or accusing him of "lying"	25.00	.63 (1.43)
Pointing out inconsistencies in the suspect's story	25.00	.53 (1.16)
Directly accusing the suspect of the theft	15.63	.25 (.76)
Making expressions of disbelief toward the suspect	12.50	.19 (.47)
Interrupting the suspect's denials	3.13	.03 (.18)
Maximization	**59.38**	**2.09 (2.61)**
Threatening the suspect with consequences	56.25	1.94 (2.48)
Exaggerating the seriousness of the offense	15.63	.16 (.37)
Leniency	**62.50**	**2.66 (3.82)**
Stating minimization themes that imply leniency	53.13	1.44 (2.47)
Making an explicit offer of leniency for confession	46.88	1.22 (1.72)
False evidence	**84.38**	**1.56 (1.02)** [*]
Bluffing about future evidence	84.38	1.47 (.98)
Lying about existing evidence	25.00	.38 (.75)
Miscellaneous tactics		
Establishing rapport via small talk	75.00	N.A.
Pressing the suspect to implicate someone else	75.00	N.A.
Encouraging the suspect to admit other illegal acts	68.75	N.A.
Appealing to the suspect's religion/conscience	21.88	N.A.
Praising or flattering the suspect	15.63	N.A.

[*] These values represent the number of different types of evidence about which police participants bluffed or lied.

arrested?"), appealing to the suspect's religion or conscience (e.g., "Some of the items had some sentimental value that's what it is. I just want to get those returned to the owner," "Are you religious at all? There is definitely right and wrong in the universe"), and the use of flattery (e.g., "You seem like a pretty good dude," "You've obviously got a great relationship with your son, great relationship with your mom").

Next we coded for the presence or absence of five discernible *police behaviors* (gathering personal information from suspects, showing crime scene photos to the suspect, asking the suspect to empty his pockets, issuing a *Miranda* warning, and obtaining a *Miranda* waiver; see Table 2). We also coded for the frequency of four discernible *suspect behaviors* (denials of guilt, denials of crime-relevant knowledge, disclosures of crime-relevant knowledge, and self-incriminating admissions; see Table 2).

For tactics and behaviors that were coded merely as present or absent (i.e., false evidence tactics, miscellaneous tactics, and police behaviors), coders exhibited an overall agreement rate of 91.67%, $\kappa = 0.83$ [95% CI: 0.77, 0.89], $p < .001$ (within each category, all $\kappa s > .69$, $p s < .001$). For those that were coded as frequency counts (i.e., confrontation, maximization, and leniency tactics, and suspect behaviors), the overall intraclass correlation (ICC) for our two raters was near perfect, ICC = .97 [95% CI: .96, .98], $p < .001$ (within each category, all ICCs > .94, $p s < .001$). Whatever disagreements or discrepancies that remained between coders were resolved via discussion.

Police reports. Two additional independent coders read and coded police reports for disclosures of these same tactics and behaviors, allowing us to compare the contents of the interrogation transcripts against the corresponding police accounts of those same interrogations. For reports, all tactics and behaviors were coded as either present (i.e., the police participant described the use of a given tactic or the presence of a given behavior) or absent (i.e., the given tactic or behavior was not noted in the police report). These coders exhibited an overall agreement rate of 97.10%, $\kappa = 0.89$ [95% CI: 0.85, 0.93], $p < .001$ (within each category, all $\kappa s > .65$, $p s < .001$). Once again, disagreements were resolved by discussion.

Table 2
Presence and Frequency of Five Police and Four Suspect Behaviors Across 32 Interrogations

Behaviors coded	% Sessions in which behavior occurred	M (SD) of occurrences
Police behaviors		
Obtaining suspect's personal information	90.63	N.A.
Showing crime scene photos to suspect	37.50	N.A.
Asking suspect to empty his pockets	18.75	N.A.
Issuing a *Miranda* warning	6.25	N.A.
Obtaining a *Miranda* waiver	6.25	N.A.
Suspect behaviors		
Disclosing crime-relevant knowledge	96.88	16.00 (9.60)
Denying guilt	90.63	7.09 (5.28)
Denying crime-relevant knowledge	75.00	2.66 (3.31)
Making an incriminating admission	12.50	.31 (1.12)

Phase 1 Results

Length of Interrogations

On average, interrogations lasted for 16.41 min ($SD = 5.50$; Range: 3.32 – 26.17) and contained 2,438.31 words ($SD = 984.11$; Range: 866 – 4,481). Our suspicion manipulation had no effect on the duration, $t(15) = 0.29$, $p = .780$, $d = 0.07$ [95% CI: −0.49, 0.63], or word count, $t(15) = 0.51$, $p = .619$, $d = 0.13$ [95% CI: −0.43, 0.68], of interrogations (note that all CIs reported here and thereafter pertain to effect sizes). On average, interrogations of the second suspect were longer in word count, $t(15) = 2.22$, $p = .042$, $d = 0.55$ [95% CI: 0.00, 1.11], but not in duration, $t(15) = 1.64$, $p = .121$, $d = 0.41$ [95% CI: −0.15, 0.97]. On average, 56.60% of all words were spoken by the interrogator ($SD = 11.63\%$; Range: 28.05–74.66%), and neither Suspicion, $t(15) = 0.56$, $p = .586$, $d = 0.14$ [95% CI: −0.42, 0.70], nor Order, $t(15) = 0.87$, $p = .401$, $d = 0.22$ [95% CI: −0.34, 0.77] affected this percentage.

No significant order effects were found on any other dependent measures, including responses to the postinterrogation questionnaires and the presence and frequency of all coded tactics and behaviors, with one exception: Suspects who were interrogated second offered more denials of crime-relevant knowledge ($M = 3.81$, $SD = 4.04$) than those who were interrogated first ($M = 1.50$, $SD = 1.86$), $t(15) = 2.21$, $p = .043$, $d = 0.55$ [95% CI: −0.01, 1.11]. Consequently, all data were collapsed across order for all subsequent analyses.

Postinterrogation Questionnaire

A 2 (Suspicion: Present vs. Absent) × 2 (Source: Police vs. Suspect) repeated-measures MANOVA was performed on the nine continuous items on the postinterrogation questionnaire (including guilt-confidence composite scores) to test for discrepancies between suspect and police perceptions of the interrogation. A multivariate effect of Source emerged, Wilks' $\Lambda = .10$, $F(9, 5) = 5.16$, $p = .043$, with significant differences on four items. Neither the multivariate effect of Suspicion, Wilks' $\Lambda = .18$, $F(9, 5) = 2.46$, $p = .167$, nor the Source × Suspicion interaction, Wilks' $\Lambda = .23$, $F(9, 5) = 1.86$, $p = .257$, was significant.

Follow-up univariate ANOVAs were performed on the four items that differed as a function of Source (see Figure 1). These analyses indicated that suspects were more confident that they would be judged as innocent ($M = 4.07$, $SD = 7.04$) than police were confident in their innocence ($M = 0.54$, $SD = 6.95$), $F(1, 13) = 4.91$, $p = .045$, $d = 0.35$ [95% CI: −0.07, 0.77]; suspects rated their denials as more credible ($M = 8.25$, $SD = 1.80$) than police rated their denials ($M = 5.61$, $SD = 2.17$), $F(1, 13) = 26.92$, $p < .001$, $d = 0.88$ [95% CI: 0.46, 1.30]; and suspects rated themselves as more cooperative ($M = 9.11$, $SD = 1.37$) than police rated them ($M = 6.89$, $SD = 2.48$), $F(1, 13) = 24.21$, $p < .001$, $d = 0.84$ [95% CI: 0.42, 1.26]. In addition, police rated the interrogations as more stressful for suspects ($M = 5.07$, $SD = 2.28$) than suspects rated it for themselves ($M = 3.46$, $SD = 2.15$), $F(1, 13) = 7.57$, $p = .016$, $d = 0.63$ [95% CI: 0.21, 1.05]. The remaining four items did not differ by Source. Police and suspects, respectively, rated police as friendly ($M s = 6.93$ & 6.86, $SD s = 1.88$ and 2.19) and mod-
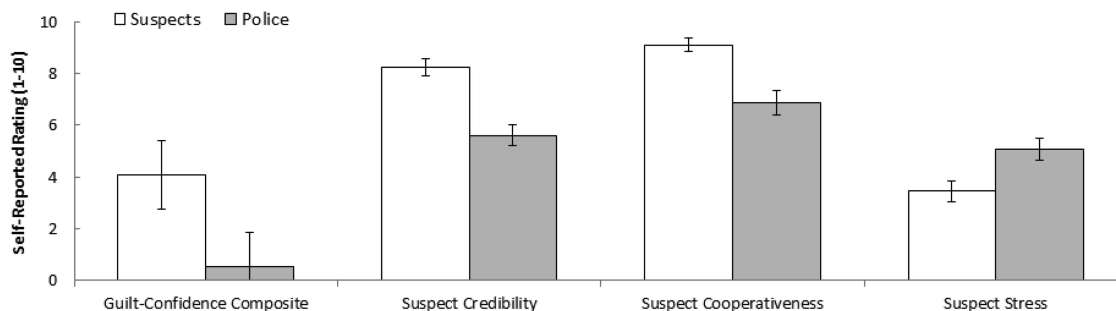
Figure 1. Effects of source (suspect vs. police) on four (of nine) continuous self-report items (Phase 1). *Note.* Higher scores on the composite measure indicate greater confidence in the suspect's projected innocence.

erately confrontational (Ms = 4.86 and 4.32, SDs = 2.26 and 2.60), and suspects as moderately anxious (Ms = 5.18 and 4.32, SDs = 2.28 and 2.36) and knowledgeable about the crime (Ms = 4.82 and 5.61, SDs = 2.79 and 2.86).

We separately analyzed three dichotomous items on the postinterrogation questionnaire. First, police judged 13 of the 32 suspects as guilty (40.62%). As with other dependent measures, suspicion did not affect these judgments, McNemar $\chi^2(1)$ = 0.82, p = .549, nor did it affect suspects' predictions of whether they would be judged guilty, McNemar $\chi^2(1)$ = 0.11, p = 1.00. Police judgments of the suspect's guilt were unrelated to suspects' predictions of whether they would be judged as guilty, McNemar $\chi^2(1)$ = 0.60, p = .607. Second, police indicated that 13 of the 32 suspects (40.62%) made suspicious remarks during the interrogation. Once again, suspicion did not affect police responses to this item, McNemar $\chi^2(1)$ = 0.14, p = 1.00 (note that three "suspicious remarks" suspects were judged innocent; three others not seen as having made suspicious remarks were judged guilty.). Police and suspect responses to this item were also unrelated to each other, McNemar $\chi^2(1)$ = 0.25, p = .804. Third, only one suspect was judged by police as having made an admission of guilt; no suspects reported having admitted guilt.

Interrogation Tactics and Behaviors

Frequencies and descriptive statistics for all 16 coded interrogation tactics and five tactic categories are shown in Table 1. Out of 32 interrogations, 56.25% featured the use of one or more confrontation tactics, 59.38% featured one or more maximization tactics, 62.50% featured one or more leniency tactics, and 84.38% featured one or more false evidence tactics. The average interrogation featured 1.63 uses of confrontation (SD = 1.98), 2.09 uses of maximization (SD = 2.61), and 2.66 uses of leniency (SD = 3.82).

With respect to presentations of false evidence, 84.38% of interrogations featured bluffs about evidence, and the average interrogation included bluffs about 1.47 different types of evidence (SD = 0.98). Police most often bluffed about having surveillance footage (75% of interrogations), followed by fingerprints (46.88%), DNA (18.75%), and eyewitnesses (3.13%). One quarter of interrogations featured outright lies about existing evidence, and the average interrogation included lies about 0.38 different types of evidence (SD = 0.75). Police most often lied about incriminating surveillance footage

(21.88% of interrogations), followed by DNA (9.38%), fingerprints (3.13%), and eyewitnesses (3.13%).

Descriptive statistics for the five coded police behaviors and four suspect behaviors are shown in Table 2. Most police solicited personal information from the suspect (90.63%); fewer showed crime scene photos (37.50%), asked the suspect to empty his pockets (18.75%), or read *Miranda* warnings (6.25%). Virtually all suspects (96.88%) disclosed some crime-relevant knowledge during the interrogation. On at least one occasion, most suspects denied guilt (90.63%) and denied crime-relevant knowledge (75.00%); relatively few suspects made incriminating statements (12.50%).

Suspicion Manipulation

Suspicion did not affect any of the coded tactics, namely, whether the interrogator exhibited one or more uses of confrontation, maximization, leniency, or false evidence, all McNemar χ^2s ≤ 1.00, ps > .60; the number of uses of confrontation, maximization, or leniency tactics, all ts < 1, ps > .75, ds < 0.10; or the number of different bluffs, $t(15)$ = 0.59, p = .566, d = 0.15 [95% CI: −0.41, 0.70], or lies, $t(15)$ = 0.19, p = .849, d = 0.05 [95% CI: −0.51, 0.61]. Suspicion likewise did not influence the likelihood of any of the five coded police behaviors, all McNemar χ^2 ≤ 2.00, ps > .50, the likelihood of any of the four coded suspect behaviors, all McNemar χ^2 ≤ 0.50, ps > .70, or the number of times that any of the four suspect behaviors occurred, all ts ≤ 1.05, ps > .30, ds ≤ 0.15.

Comparison of Transcripts and Reports

In comparison to the interrogation transcripts, which contained an average of 2,438.31 words (SD = 984.11; Range: 866–4,481), police reports contained an average of 1,224.44 words (SD = 365.52; Range: 570–1,875). Descriptions of the crime scene contained fewer words (M = 258.06, SD = 76.31) than descriptions of either the first (M = 452.44, SD = 174.94) or the second (M = 478.31, SD = 183.06) interrogation, $F(2, 30)$ = 23.01, p < .001, η_p^2 = .61, which did not differ in word count.

We sought to measure both the accuracy and completeness of police reports by noting errors of commission and omission in relation to the coded transcripts. For this purpose, we compared the

presence of all coded tactics and behaviors in the interrogation transcripts against the presence of those same tactics and behaviors in the corresponding police reports. Errors of commission were extremely rare. Across 25 coded tactics and behaviors in 32 interrogations (a total of 800 coded details), only four errors of commission were found (0.50%). One interrogator mistakenly reported having lied about evidence to one suspect and having shown crime scene photos to the other; a second interrogator mistakenly reported using minimization tactics; and a third mistakenly reported having offered leniency in exchange for a confession.

In contrast, errors of omission were prevalent. Figure 2 shows the frequencies of usage and reporting for all coded tactics and behaviors. As noted earlier, a majority of interrogations included the use of one or more tactics involving confrontation (56.25%), maximization (59.38%), leniency (62.50%), and false evidence (84.38%). When used, however, the corresponding reports described one or more of these tactics only 22.22%, 15.79%, 40%, and 66.67% of the time, respectively (e.g., in 77.78% of reports of interrogations in which confrontation was used one or more times, no mention of any confrontation tactic was made). In short, a

number of different tactics were frequently used during interrogations but did not appear in corresponding police reports.

With respect to suspect behaviors, virtually all suspects (96.88%) disclosed some crime-relevant knowledge regarding the mock theft, and these disclosures were almost always reported by police (96.77%). Most suspects also issued denials of guilt (90.63%), which were also typically reported (89.66%). However, while 75% of suspects denied crime-relevant knowledge at least once, only 33.33% of the corresponding reports noted such denials (in contrast, three out of four suspects who made incriminating admissions were reported to have done so).

To sum up: Two sets of findings emerged from Phase 1 of our study. First, police and suspects diverged in their perceptions of the interrogations, with suspects seeing themselves as more credible in their denials, more cooperative, and more confident in their presentation of innocence. Second, a comparison of interrogation transcripts and police reports indicated that errors of omission were prevalent, with police underreporting the use of various tactics. In light of these results, Phase 2 assessed whether observers' impressions of the interrogations were differently influenced
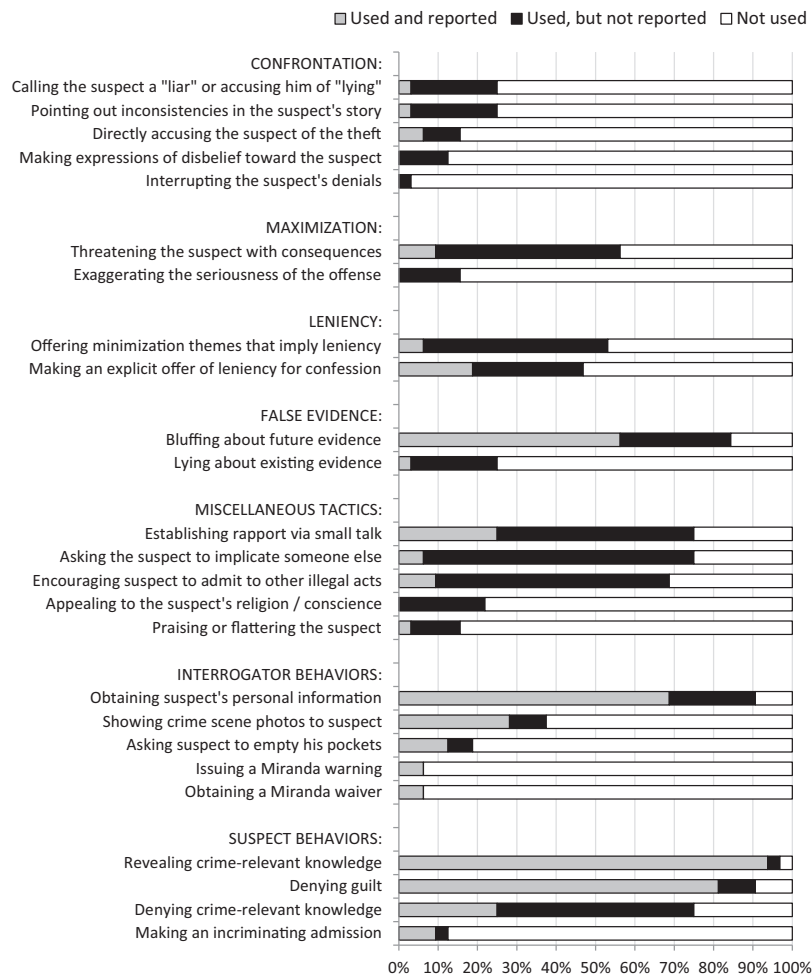

Figure 2.   Occurrence and reporting of interrogation tactics, police behaviors, and suspect behaviors across 32 interrogations and corresponding police reports.

by the police reports that were generated relative to full and objective records.

Phase 2 Method

Participants and Design

Participants in Phase 2 were 96 undergraduates, 64 of whom were female, with a mean age of 19.45 (*SD* = 2.84). A total of 38.54% self-identified as Hispanic, 17.71% as White, 15.63% as Asian, 12.50% as Black, and 15.63% as multiracial or Other.

To determine whether outside fact finders would form different or more accurate impressions when they had access to verbatim transcripts of suspect interrogations, compared to secondhand police reports, participants read a brief summary of the case and were randomly assigned to either read a police report or an interrogation transcript of one of the 32 Phase 1 interrogations. These conditions enabled us to compare directly the two alternative written accounts of the process. Next we sought to assess what impressions participants would form if they also listened to audiotapes of these same sessions, thereby having access to vocal and paralinguistic cues emanating from both police and suspect participants (e.g., prosodics such as pitch, volume, stress, tempo, pitch, rhythm, pauses, points of stress, and intonation—the kinds of cues that often lead laypeople to draw erroneous inferences of truth and deception; see DePaulo et al., 2003; Bond & DePaulo, 2006). Thus, a third group of participants—equipped with headphones and an MP3 player—listened to the audio recording while following along on the transcript. This group enabled us to compare the two verbatim conditions—one written, the second accompanied by audio cues. Because no significant effects of manipulated suspicion were obtained in Phase 1, this second experiment combined rather than distinguished these two groups of suspects in all analyses.

Procedure

Participants completed the study in sessions that included from four to eight individuals. All participants in a given session were in the same experimental condition, but they read and/or listened via headphones to different Phase 1 interrogations. After giving their informed consent and providing basic demographic information (i.e., age, gender, and race), all participants read the same one-page background summary of the crime and investigation that occurred during Phase 1. The summary described the theft of an unknown amount of cash from a briefcase inside a police station, described the crime scene and the contents of the briefcase, and explained that two suspects were identified who were known to have been in the vicinity of the briefcase around the time of the theft. Participants were told that they would now be given an account of the interrogation of one of these two suspects. Participants then read the police report (*report* condition) or the full transcript (*transcript* condition), or they listened to an audio recording (*audio* condition) of one of the 32 Phase 1 interrogations. When they were finished, they completed a questionnaire that measured their perceptions of the interrogator and suspect, after which they were fully debriefed.

Dependent Measures

The questionnaire administered to Phase 2 participants contained 13 items that paralleled those answered by police and suspects in Phase 1. First, participants indicated whether they believed the suspect to be guilty or innocent and rated their confidence in that judgment on a scale from 1 (*not at all*) to 10 (*very*). As in Phase 1, these items were combined to form a guilt-confidence composite score that could range from −10 (*highly confident guilty judgment*) to +10 (*highly confident innocent judgment*). Participants also gave continuous ratings of how credible the suspects' denials were, how knowledgeable the suspect was about the crime, how cooperative the suspect was, how hard the interrogator tried to get the suspect to confess, how friendly the interrogator was, how confrontational the interrogator was, how anxious the suspect was, and how stressful they thought the interrogation was for the suspect. One new continuous item was added for Phase 2, which asked participants to rate how much pressure the interrogator placed on the suspect during the interrogation. These nine continuous ratings were each given on a scale from 1 (*not at all*) to 10 (*very*). Participants also gave dichotomous yes/no judgments as to whether the suspect had made any suspicious remarks and/or any admissions of guilt during the interrogation. Four participants (4.17%) neglected to answer one or more items, so their data are missing from the relevant analyses.

Phase 2 Results

Two a priori sets of comparisons framed our analyses. First and foremost, we sought to compare the two written accounts of the interrogations—secondhand police reports versus verbatim transcripts. Next we compared the two verbatim accounts—written transcripts alone versus accompanied by an audiotape.

Written Police Reports Versus Transcripts

Our most important prediction was that reading transcripts of interrogations would improve fact-finding accuracy relative to reading the accounts contained in police incident reports by rendering participants less likely to judge innocent suspects as guilty. Supporting this prediction, participants in the Transcript condition misjudged the suspect as guilty less often than did those in the Report condition (9.38% vs. 31.25%, respectively), $\chi^2(1) = 4.73$, $p = .030$, $\varphi = .27$, $OR = 4.39$ [95% CI: 1.08, 17.86].

We then compared the Transcript and Report conditions in terms of their continuous ratings of the interrogator and suspect on our questionnaire. A one-way MANOVA on these 10 items did not reach a conventional level of significance, Wilks' $\Lambda = .74$, $F(10, 47) = 1.65$, $p = .123$, but it did reveal a large effect size ($\eta_p^2 = .26$) and suboptimal power ($1\text{-}\beta = .71$). In light of this, and to more precisely focus on our most pertinent system-relevant measures, we performed univariate t tests designed to directly test our prediction that presentation medium would in particular impact perceptions of *guilt* and *coercion*.

Significant univariate differences were found on three items, each of which showed a medium to large effect size (see Figure 3; Cohen, 1988). Consistent with the aforementioned binary judgments, participants who read a transcript were more confident in the suspect's innocence (*M* = 5.86, *SD* = 4.82) than were those who read the corresponding police report (*M* = 2.34, *SD* = 7.12), $t(56) = 2.20$, $p = .032$, $d = 0.59$ [95% CI: −0.95, 2.14]. Importantly as well, compared to those in the Report condition, participants who read a transcript also rated the interrogator as having
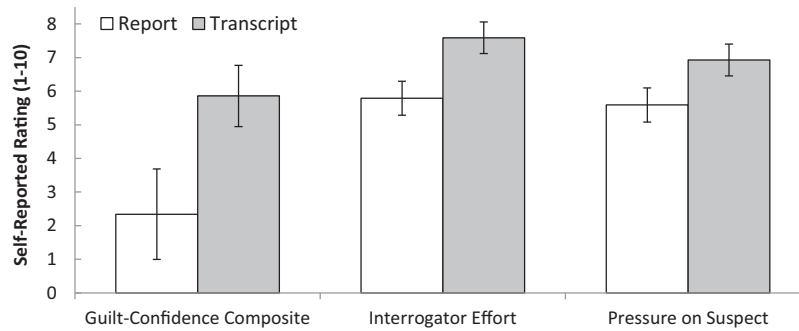
Figure 3. Comparison of written transcripts and reports on Phase 2 perceptions of interrogators and suspects.
Note. Higher scores on the composite measure indicate greater confidence in the suspect's innocence.

exerted more pressure on the suspect (Ms = 6.93 and 5.59, SDs = 2.36 and 2.43), $t(56)$ = 2.14, p = .037, d = 0.57 [95% CI: 0.03, 1.18], and as having tried harder to obtain a confession (Ms = 7.59 and 5.79, SDs = 2.50 and 2.70), $t(56)$ = 2.62, p = .011, d = 0.70 [95% CI: 0.04, 1.36].

No differences were found on the other continuous ratings, all ts < 1.15, ps > .26. Overall, participants in the Transcript and Report conditions, respectively, rated the interrogation as moderately stressful (Ms = 5.28 and 5.66, SDs = 2.83 and 3.11) and the suspect as moderately anxious (Ms = 5.24 and 5.48, SDs = 2.69 and 2.89), credible (Ms = 6.72 and 6.21, SDs = 1.79 and 2.32), and knowledgeable (Ms = 5.48 and 5.38, SDs = 2.52 and 2.62)—and as highly cooperative (Ms = 8.52 and 8.14, SDs = 1.38 and 2.01). Transcript and Report participants also rated the interrogator both as somewhat friendly (Ms = 6.07 and 5.31, SDs = 2.88 and 2.27) and somewhat confrontational (Ms = 6.07 and 5.41, SDs = 2.30 and 2.15).

Two additional questionnaire items asked whether suspects had made any suspicious remarks or admissions of guilt. Participants in the Transcript condition were no more or less likely than those in the Report condition to believe that the suspect had made suspicious remarks (37.50% vs. 28.13%, respectively), $\chi^2(1)$ = 0.64, p = .424, φ = .10, OR = 1.53 [95% CI: 0.54, 4.39], nor were they more or less likely to believe that he had admitted guilt (13.79% vs. 6.90%, respectively), $\chi^2(1)$ = 0.74, p = .389, φ = .11, OR = 2.16 [95% CI: 0.36, 12.84].

Verbatim Transcripts Versus Audiotapes

Next, we compared the judgments of participants in the Transcript and Audio conditions who received the same verbal content but presented in a different modality (in writing vs. audio recording). Interestingly, we found that participants in the Audio condition misjudged the suspect as guilty more often than did those in the Transcript condition (37.50% vs. 9.38%, respectively), $\chi^2(1)$ = 7.05, p = .008, φ = .33, OR = 5.80 [95% CI: 1.45, 23.23], a difference that is not surprising in light of research showing that people are not intuitively accurate judges of truth and deception. Audio participants were not, however, more likely to believe that their suspect had made suspicious remarks (53.13% vs. 37.50%, respectively), $\chi^2(1)$ = 1.58, p = .209, φ = .16, OR = 1.89 [95% CI: 0.70, 5.12], or that he had made an admission of guilt (12.90% in both conditions), $\chi^2(1)$ = 0.00, p = 1.00, φ = .00, OR = 1.00 [95% CI: 0.23, 4.42].

A one-way MANOVA on continuous ratings of the interrogator and suspect did not reach significance, Wilks' Λ = .73, $F(10, 51)$ = 1.85, p = .075. Once again, however, this analysis revealed a large effect size (η_p^2 = .27) and a relative lack of power (1-β = .78). Follow-up t tests on the most relevant continuous items revealed two significant univariate differences with medium to large effect sizes. First, participants in the Audio condition were less confident in the suspect's innocence (M = 1.19, SD = 7.55) than were those in the Transcript condition (M = 5.97, SD = 4.68), $t(60)$ = 2.99, p = .004, d = 0.77 [95% CI: −0.77, 2.31]. Second, participants in the Audio condition rated the interrogator as having exerted less pressure on the suspect (M = 5.39, SD = 3.13) than did those in the Transcript condition (M = 6.89, SD = 2.46), $t(60)$ = 2.10, p = .040, d = 0.54 [95% CI: −0.15, 1.23]. No significant differences were found on the remaining items, all ts < 1.87, ps > .06.

Discussion

In light of the problems associated with confession evidence, numerous social scientists, legal scholars, and practitioners have recommended a policy reform that would require the electronic recording of entire suspect interviews and interrogations—not just the resulting confessions (e.g., American Psychology-Law Society white paper—Kassin et al., 2010). Historically, such a requirement has proved controversial, drawing opponents from the law enforcement community who have speculated about the possible adverse effects on police, suspects, and juries (e.g., Inbau et al., 2001; for an overview of opposition arguments see Sullivan, 2008; Thurlow, 2005).

The present research was designed to test for an important possible benefit of audio recording full interrogations: that these recordings provide a more accurate factual account of the interrogation behavior of police and suspects than would otherwise be derived from police reports. Indeed, a common argument for recording interrogations is that it is the most effective way to memorialize the process by which a statement was taken and, hence, increase the fact-finding accuracy of judges and juries. In current practice, whereby detectives produce reports from contemporaneous or retrospective notes, disputes often arise as to a number of issues concerning the behavior of police and suspects—most notably, whether the police Mirandized suspects in a timely manner, whether they used certain coercive tactics, and whether

the details contained within a confession originated with the suspect or came about through a process of contamination.

Police notes from interviews and interrogations are routinely accepted in the courts in lieu of an objective recording. Yet the accuracy and impact of these notes has never been tested in the context of a suspect interrogation. In light of the importance of the policy in question as well as basic research on memory for conversations, we conducted a two-phased experiment. In Phase 1, experienced police officers investigated a mock crime scene, interviewed two innocent suspects—one of whom was presented as suspicious on the basis of his behavior—and then filed an incident report. All sessions were covertly audio recorded; these recordings were later used to assess the accuracy of the reports. In Phase 2, lay participants read either the police report or the interrogation transcript, or they listened to an audiotape. Our goal was to compare the accuracy of these "fact finders" as a function of the information they were provided.

In Phase 1, two notable sets of results were obtained. First, in parallel postinterrogation questionnaires, police and their suspects diverged in their perceptions of the suspect's behavior. Specifically, suspects believed that they would be seen as more innocent than the police actually believed them to be; they also rated their denials as more credible, their stress levels as lower, and their behavior as more cooperative. Along with the finding that police judged 41% of suspects as guilty, and has having made suspicious remarks, this result demonstrates a phenomenon often seen in real trials: That detectives and defendants often testify disparately as to what transpired during the process of interrogation.

The second key result emerged from the comparison between the tactics that police actually used during their interrogations, as later coded from the tapes, and the tactics police said they used in their incident reports. Consistent with past research using a mock crime paradigm (e.g., Kassin et al., 2014), in naturalistic field settings involving real interrogations (e.g., Feld, 2013; Leo, 1996), and in self-report surveys of police (e.g., Kassin et al., 2007), we found that police participants commonly used confrontation, maximization, minimization and leniency, and presentations of false evidence in the form of bluffing and outright lies—even in the context of interrogations that were brief and without high-stakes consequences.

In this regard, two important points are worth noting. First, closely replicating the mock interrogation results reported by Kassin et al. (2014), the police participants in our study used the false evidence ploy at a very high rate. This result contrasts with more realistic field-based data (e.g., Leo, 1996) and self-report surveys (e.g., Kassin et al., 2007) indicating that this tactic is used more sparingly. Second, when the usage frequencies were compared with the incident reports, even though police participants were instructed to state as close to verbatim as possible what tactics they used, numerous errors of omission were observed. Specifically, whereas most interrogations involved the use of confrontation, maximization, leniency, and false evidence at rates of 56%, 59%, 63%, and 84%, respectively, corresponding reports described these tactics only 22%, 16%, 40%, and 67% of the time, respectively. In actual cases, such underreporting, whether purposeful or inadvertent, could influence judges who rule on the voluntariness of statements taken and juries who rule on their credibility.

In Phase 2, we sought primarily to determine whether outside fact finders would form different or more accurate impressions when they had access to written transcripts of suspect interviews compared to the written police reports. On the most important dependent measure, perceptions of the suspect's guilt or innocence, the results confirmed expectations: Participants were significantly less likely to misjudge the innocent suspect as guilty when they read a written transcript of the interrogation than when they read the corresponding police report (9% vs. 31%). Moreover, participants who read a police report, compared to those who read the full transcript, believed that police officers had applied less pressure on the suspect and tried less hard to get the suspect to confess. In short, observers whose information basis was one of the police reports relative to one of the transcripts, saw the process as less pressure-filled and the innocent suspect as more guilty.

In addition to comparing the two alternative forms of written information—police reports and transcripts—we wondered what impressions observers would have if they listened to an interrogation audiotape. On the one hand, this richer medium of presentation presents exactly the same verbal content along with the transcripts—perhaps, therefore, with the same result. On the other hand, audiotapes add complex vocal and paralinguistic cues from both police and suspect participants—the kinds of cues that often lead laypeople to draw unwarranted and erroneous inferences of truth and deception (e.g., DePaulo et al., 2003; Bond & DePaulo, 2006). Suggesting that the latter cues influenced perceptions, apart from verbal content, several results showed that participants who heard an audiotape, relative to those who merely read a transcript, fell prey to some of these effects. Most notably, for example, 38% misjudged the suspect to be guilty, a number that was significantly higher than in the transcript-only condition.

As a matter of policy, one could argue from our results that perhaps strictly content-focused interrogation transcripts—without access to audio and visual cues—would provide a sufficient basis for fact finding. Such a conclusion would not be warranted. While a transcript fully communicates the verbal text of a police–suspect interaction, it does not depict potentially important aspects of the suspect (e.g., his or her physical condition, appearance, attire, voice, and demeanor; whether he or she is seated in a corner or handcuffed) or the police officers (e.g., their number, size, and proximity to the suspect; whether they are uniformed or in plain clothes; whether weapons are visible; whether they raise their voices). In this regard, extensive research indicates that fact finders render more balanced and accurate judgments from "equal focus" video recordings that show both the suspect and police rather than one or the other (Lassiter, Diamond, Schmidt, & Elek, 2007; Lassiter, Geers, Handley, Weiland, & Munhall, 2002).

Summary, Implications, and Limitations

The most important signal to emerge from our study is one that strongly supports a requirement that all suspect interviews and interrogations be recorded and transcribed in order to provide a more accurate account of the process and improve the fact-finding performance of judges and juries. This signal is embodied in the following main findings: (a) Police and suspects diverged in their perceptions of the suspect's behavior during the interrogation sessions in which both parties were present; (b) Police committed frequent errors of omission in their Incident Reports, underreporting their use of confrontation, maximization, leniency, and presentations of false evidence; and (c) Phase 2 participants who read a police report, compared to those who read a full verbatim

transcript, perceived the process as less pressure-filled and were more likely to misjudge innocent suspects as guilty.

Taken as a whole, these findings help to explain the second problem with false confessions (the first being that they occur): That they are too often believed by judges and juries equipped only with secondhand information about the process by which the statements were elicited (Kassin, 2012). On this issue, results from Phase 2 suggest two important points. First, participant observers were aided by having access to the more accurate content provided by verbatim transcripts, suggesting, perhaps, this previously neglected potential benefit of electronic recording. Second, participant observers did not benefit from the addition of audio recordings of the interrogations, leaving open the empirical question, untested in our study, of whether additional access to equal-focus video would improve fact-finding performance (for a discussion, see Snyder, Lassiter, Lindberg, & Pinegar, 2009).

One might wonder whether the errors of omission we observed, in the form of underreporting the use of interrogation tactics, were purposeful or inadvertent. As a strictly empirical matter, our data do not permit us to weigh in on this issue. However, given the context (i.e., these errors were observed in a mock-crime-and-investigation study; police participants wrote Incident Reports that would not later be shown to a prosecutor, entered into evidence at trial, or become the subject of sworn testimony), common sense would suggest that these errors were inadvertent. Such inadvertence is often seen in the contamination of false confessions with accurate crime details, a phenomenon observed in actual cases (to illustrate, see Trainum, 2007; for reviews, see Garrett, 2010, 2015; also see Lamb et al., 2000). In this regard, it is important to note that we did not test for—and our estimates do not account for the possibility of—conscious or motivated failures of police to recall or report certain aspects of the interrogations they conduct. Because we asked for all reports to be submitted within 48 hours, we also did not test for the possibility that reporting errors would increase as a function of longer intervals between interrogations and reports.

The present research is potentially limited in important ways. Our police participants were trained and seasoned professionals, ranging in age from 26 to 55 and having an average of 16 years of law enforcement experience, which included numerous suspect interviews. In the context of our mock-crime-and-investigation paradigm, however, these participants were limited to 5 minutes for crime scene analysis and two relatively brief interrogations— each lasting an average of only 17 min, and a 48-hr time limit for submission of reports. Whether our results would generalize to longer and more consequential investigations remains an important empirical question for further research. On the one hand, one might argue that our police participants were not as motivated to recall the details of their mock interrogations as they are in actual practice, causing us to overestimate errors of omission and their effects. On the other hand, one might argue from a cognitive perspective that because it is so much easier to recall brief conversations that last only a few minutes, rather than hours-long interrogations, our results underestimate the potential for errors of omission and do not adequately test for possible errors of commission, including source attribution errors.

Our results are also limited by the fact that our manipulation of prior suspicion had no significant effects whatsoever—not on the length of interrogations in minutes or words, participants' perceptions of the exchange, police perceptions of the suspect's guilt, or the coded

interrogation tactics that were used. We had hoped to assess whether the accuracy of police reports and the impressions they elicited in observers were moderated by investigators' preexisting degree of suspicion. In prior research, suspicion was varied by the presentation of a base rate (i.e., 80 vs. 20% of suspects in this study are guilty of the mock crime; see Kassin, Goldstein, & Savitsky, 2003). In this study, however, because each investigator was set to conduct two interrogations, we manipulated relative suspiciousness by describing the demeanor of the two suspects—one as calm, the other as anxious and evasive. Lacking a manipulation check to ensure that investigators drew the intended inferences from the experimenter's description of the two suspects, we cannot adequately evaluate the possible effects of suspicion on accuracy and bias in police reports. Testing this confirmation bias hypothesis thus remains an important avenue for follow-up research.

We should also comment on the mock-crime-and-investigation paradigm we used, modeled after that previously reported in Kassin et al. (2014). Conducted in a police station and involving a collaboration of experienced law enforcement participants, this experiment contained a high level of ecological validity. As noted by Kassin et al. (2014), however, such data are difficult to collect during the workday from on-duty and off-duty officers, detectives, and sergeants. For that reason, our sample was smaller than we had hoped it would be, thereby limiting the power of our study and hence our ability to analyze for individual differences in experience or training among police participants.

One final limitation concerns the medium through which Phase 2 participants, mimicking fact finders, made their judgments. Specifically, they read a brief description of the crime followed by the "raw data" of a police report or verbatim transcript or audio recording of an interrogation. Police participants who had interrogated mock suspects did not testify as to their experience—and they were not cross examined. Although research suggests that the process of cross examination may have variable effects—for example, helping jurors to become more discerning of scientific experts (Austin & Kovera, 2015), yet impairing an eyewitness's memory (Valentine & Maras, 2011)—it is nevertheless the natural process through which fact finders are informed about interviews and interrogations. More research is needed to determine if cross-examination serves to correct for the underreporting of tactics used.

With an accumulation of DNA exonerations illuminating the problem of false confessions, and with research indicating the dual risk that innocent people might confess to crimes they did not commit and that judges and juries may well believe these false confessions, it is easy to understand calls to reform that focus on the recording of interrogations. Limitations notwithstanding, the present study adds to a growing empirical literature indicating the need for such a requirement, if only to ensure the accuracy and completeness of memorial accounts of key transactions between police and their suspects—accounts that form the basis of decisions routinely made by judges and juries.

References

American Bar Association. (2004). *Resolution 8A—Videotaping custodial interrogations*. Approved February, 9, 2004.

American Psychological Association. (2014). *Resolution on interrogations of criminal suspects*. Retrieved from http://www.apa.org/about/policy/interrogations.aspx

Appleby, S. C., Hasel, L. E., & Kassin, S. M. (2013). Police-induced confessions: An empirical analysis of their content and impact. *Psychology, Crime & Law, 19,* 111–128. http://dx.doi.org/10.1080/1068316X.2011.613389

Austin, J. L., & Kovera, M. B. (2015). Cross-examination educates jurors about missing control groups in scientific evidence. *Psychology, Public Policy, and Law, 21,* 252–264. http://dx.doi.org/10.1037/law0000049

Boetig, B. P., Vinson, D. M., & Weidel, B. R. (2006). Revealing incommunicado. *FBI Law Enforcement Bulletin, 75,* 1–8.

Bond, C. F., Jr., & DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality and Social Psychology Review, 10,* 214–234. http://dx.doi.org/10.1207/s15327957pspr1003_2

Bruck, M., Ceci, S. J., & Francoeur, E. (1999). Accuracy of mothers' memories of conversations with their preschool children. *Journal of Experimental Psychology: Applied, 5,* 89–106. http://dx.doi.org/10.1037/1076-898X.5.1.89

Buckley, D. M., & Jayne, B. C. (2005). *Electronic recording of interrogations.* Eagle River, WI: Hahn Printing,.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Deffenbacher, K. A., Bornstein, B. H., Penrod, S. D., & McGorty, E. K. (2004). A meta-analytic review of the effects of high stress on eyewitness memory. *Law and Human Behavior, 28,* 687–706. http://dx.doi.org/10.1007/s10979-004-0565-x

DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin, 129,* 74–118. http://dx.doi.org/10.1037/0033-2909.129.1.74

Drizin, S. A., & Reich, M. J. (2004). Heeding the lessons of history: The need for mandatory recording of police interrogations to accurately assess the reliability and voluntariness of confessions. *Drake Law Review, 52,* 619–646.

Federal Rules of Evidence. (2015). Arlington, VA: Federal Evidence Review. Retrieved from http://federalevidence.com/downloads/rules.of.evidence.pdf

Feld, B. C. (2013). *Kids, cops, and confessions: Inside the interrogation room.* New York, NY: New York University Press.

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

Garrett, B. L. (2015). Contaminated confessions revisited. *Virginia Law Review, 101,* 395–453.

Geller, W. A. (1993). *Videotaping interrogations and confessions: A preliminary examination of issues.* Washington, DC: National Institute of Justice.

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

Inbau, F., Reid, J., Buckley, J. P., & Jayne, B. C. (2001). *Criminal interrogation and confessions* (4th ed.). Gaithersburg, MD: Aspen.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114,* 3–28. http://dx.doi.org/10.1037/0033-2909.114.1.3

Kassin, S. M. (1997). The psychology of confession evidence. *American Psychologist, 52,* 221–233. http://dx.doi.org/10.1037/0003-066X.52.3.221

Kassin, S. M. (2005). On the psychology of confessions: Does innocence put innocents at risk? *American Psychologist, 60,* 215–228. http://dx.doi.org/10.1037/0003-066X.60.3.215

Kassin, S. M. (2012). Why confessions trump innocence. *American Psychologist, 67,* 431–445. http://dx.doi.org/10.1037/a0028212

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. http://dx.doi.org/10.1007/s10979-009-9188-6

Kassin, S. M., Goldstein, C. C., & Savitsky, K. (2003). Behavioral confirmation in the interrogation room: On the dangers of presuming guilt.

Law and Human Behavior, 27, 187–203. http://dx.doi.org/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,* 33–67. http://dx.doi.org/10.1111/j.1529-1006.2004.00016.x

Kassin, S. M., Kukucka, J., Lawson, V. Z., & DeCarlo, J. (2014). Does video recording alter the behavior of police during interrogation? A mock crime-and-investigation study. *Law and Human Behavior, 38,* 73–83. http://dx.doi.org/10.1037/lhb0000047

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

Kleider, H. M., Pezdek, K., Goldinger, S. D., & Kirk, A. (2008). Schema-driven source misattribution errors: Remembering the expected from a witnessed event. *Applied Cognitive Psychology, 22,* 1–20. http://dx.doi.org/10.1002/acp.1361

Lamb, M. E., Orbach, Y., Sternberg, K. J., Hershkowitz, I., & Horowitz, D. (2000). Accuracy of investigators' verbatim notes of their forensic interviews with alleged child abuse victims. *Law and Human Behavior, 24,* 699–708. http://dx.doi.org/10.1023/A:1005556404636

Lassiter, G. D., Diamond, S. S., Schmidt, H. C., & Elek, J. K. (2007). Evaluating videotaped confessions: Expertise provides no defense against the camera-perspective effect. *Psychological Science, 18,* 224–226. http://dx.doi.org/10.1111/j.1467-9280.2007.01879.x

Lassiter, G. D., Geers, A. L., Handley, I. M., Weiland, P. E., & Munhall, P. J. (2002). Videotaped interrogations and confessions: A simple change in camera perspective alters verdicts in simulated trials. *Journal of Applied Psychology, 87,* 867–874. http://dx.doi.org/10.1037/0021-9010.87.5.867

Lassiter, G. D., & Meissner, C. A. (Eds.). (2010). *Police interrogations and false confessions: Current research, practice, and policy recommendations.* Washington, DC: American Psychological Association. http://dx.doi.org/10.1037/12085-000

Leo, R. A. (1996). Inside the interrogation room. *Journal of Criminal Law and Criminology, 86,* 266–303.

MacLeod, C. M., Gopie, N., Hourihan, K. L., Neary, K. R., & Ozubko, J. D. (2010). The production effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 671–685. http://dx.doi.org/10.1037/a0018785

Morgan, C. A., III, Hazlett, G., Doran, A., Garrett, S., Hoyt, G., Thomas, P., & Southwick, S. M. (2004). Accuracy of eyewitness memory for persons encountered during exposure to highly intense stress. *International Journal of Law and Psychiatry, 27,* 265–279. http://dx.doi.org/10.1016/j.ijlp.2004.03.004

Neisser, U. (1981). John Dean's memory: A case study. *Cognition, 9,* 1–22. http://dx.doi.org/10.1016/0010-0277(81)90011-1

Schacter, D. L. (2001). *The seven sins of memory: How the mind forgets and remembers.* Boston, MA: Houghton Mifflin.

Schmidt, M. S. (2014, May 23). In policy change, justice dept. to require recording of interrogations. *The New York Times,* p. A14.

Snyder, C. J., Lassiter, G. D., Lindberg, M. J., & Pinegar, S. K. (2009). Videotaped interrogations and confessions: Does a dual-camera approach yield unbiased and accurate evaluations? *Behavioral Sciences & the Law, 27,* 451–466. http://dx.doi.org/10.1002/bsl.875

Sullivan, T. P. (2004). *Police experiences with recording custodial interrogations.* Chicago, IL: Northwestern University Law School, Center on Wrongful Convictions.

Sullivan, T. P. (2008). Recording federal custodial interviews. *The American Criminal Law Review, 45,* 1297–1345.

Sullivan, T. P., Vail, A. W., & Anderson, H. W. (2008). The case for recording police interrogation. *Litigation, 34,* 1–8.

The Justice Project. (2007). *Electronic recording of custodial interrogations: A policy review.* Washington, DC: Author.

Thurlow, M. D. (2005). Lights, camera, action: Video cameras as tools of justice. *The John Marshall Journal of Computer & Information Law, 23,* 771–813.

Trainum, J. (2007, September 20). I took a false confession–so don't tell me it doesn't happen! *The California Majority Report.* Retrieved from http://seeingtheforest.com/i-took-a-false-confession-so-dont-tell-me-it-doesnt-happen/

Valentine, T., & Maras, K. (2011). The effect of cross-examination on the accuracy of adult eyewitness testimony. *Applied Cognitive Psychology, 25,* 554–561. http://dx.doi.org/10.1002/acp.1768

Appendix

Self-Report Questionnaire

1) In your opinion, is the suspect you just interrogated guilty or innocent?

(*In your opinion, did the officer who interrogated you believe you were guilty or innocent?*)

2) How confident are you in this opinion?

(*How confident are you that this was his/her opinion?*)

3) In your opinion, how credible were (the suspect's/*your*) denials?

4) How much did this suspect seem to know about the crime?

(*In your opinion, how much did you know about the crime before you were questioned?*)

5) In your opinion, how open and cooperative (was this suspect/*were you*) in (his/*your*) attitude and demeanor?

6) Did (this suspect/*you*) make any suspicious or self-incriminating remarks?

7) How hard did (you/*the officer*) try to get (this suspect/*you*) to confess?

[Phase 2 only: How much pressure did the police officer place on the suspect during the interrogation?]

8) How friendly, sympathetic, and understanding was (your behavior toward this suspect/*the officer toward you*)?

9) How confrontational (were you toward this suspect/*was the officer toward you*)?

10) How stressful do you think the interrogation was for this suspect?

(*How stressful was this interrogation for you?*)

11) Overall, how anxious (was this suspect/*were you*) during the interrogation?

12) Did (the suspect/*you*) make any admissions or confessions of guilt?