

ORIGINAL ARTICLE

Expertise in Deception Detection Involves Actively Prompting Diagnostic Information Rather Than Passive Behavioral Observation

Timothy Roland Levine¹, David Daniel Clare², J. Pete Blair³,
Steve McCornack^{2,4}, Kelly Morrison², & Hee Sun Park¹

1 School of Media and Communication, Korea University, Seoul, 136-701, Republic of Korea

2 Department of Communication, Michigan State University, East Lansing, MI 48823, USA

3 School of Criminal Justice, Texas State University, San Marcos, TX, USA

4 Department of Communication, Michigan State University, East Lansing, MI 48824-1212, USA

In a proof-of-concept study, an expert obtained 100% deception-detection accuracy over 33 interviews. Tapes of the interactions were shown to N = 136 students who obtained 79.1% accuracy (Mdn = 83.3%, mode = 100%). The findings were replicated in a second experiment with 5 different experts who collectively conducted 89 interviews. The new experts were 97.8% accurate in cheating detection and 95.5% accurate at detecting who cheated. A sample of N = 34 students watched a random sample of 36 expert interviews and obtained 93.6% accuracy. The data suggest that experts can accurately distinguish truths from lies when they are allowed to actively question a potential liar, and nonexperts can obtain high accuracy when viewing expertly questioned senders.

doi:10.1111/hcre.12032

Two of the more intriguing findings in the deception-detection literature involve the relatively poor levels of accuracy that are nearly universally observed across prior experiments and the lack of reliable moderating effects for judge expertise. Meta-analyses of deception-detection experiments have found that people are typically only slightly better than chance at detecting deception, and that message judges with professional expertise do not perform much differently than college students (Aamodt & Custer, 2006; Bond & DePaulo, 2006).

Consider that the mean accuracy in deception detection, as reported in meta-analysis, is 53.46% (weighted correct truth–lie classification; Bond & DePaulo, 2006). This level of performance is similar to the 53.1% accuracy in which Bem's (2011, study 1) subjects were able to predict random future events. What makes this near-equivalence so remarkable is that existing theory predicts that people should be substantially better-than-chance at deception detection, whereas extant theory does not account for the better-than-chance precognition findings reported by Bem.

Corresponding author: Timothy Roland Levine; e-mail: levinet111@gmail.com

The fact that people's deception-detection accuracy is roughly equal to ability to foresee random events poses perplexing theoretical and methodological questions. The current research tests the predictions of various theoretical accounts of expert lie detection, and creates an experimental environment conducive to accurate expert lie detection.

Deception theory, judge expertise, and deception-detection accuracy

According to prominent deception theories, lies should be detectable—at least under certain favorable conditions. Ekman and Friesen's (1969) classic description of leakage and deception cues specifies that high-stakes deception produces emotional responses which, in turn, result in behaviors indicative of lying. People adept at spotting leakage and deception cues should be adept at spotting high-stakes lies. According to Zuckerman, DePaulo, and Rosenthal's (1981) four-factor theory, deception is associated with identifiable nonverbal behaviors because lies are linked with emotions, lies require greater cognitive effort than do honest messages, lying is associated with arousal, and liars over control their behaviors. Each of these deception-linked states is associated with observable behaviors which should enable skilled judges to distinguish truths from lies. Similarly, Buller and Burgoon's (1986) interpersonal deception theory posits that liars behave differently than honest senders; and that detection accuracy is a function of liar skill relative to the message judge skill. Most recently, Vrij, Granhag, and Porter (2010) and Vrij and Granhag (2012) concluded that lying is more cognitively effortful than honest communication, and they suggest that questioning strategies which increase a sender's cognitive load will enhance lie detection (for a contrasting view, see McCornack, 1997; McCornack, Morrison, Paik, Wisner, & Zhu, in press).

The theoretical specification that at least some lies are detectable by some people does not square well with the empirical observation that people are almost universally poor lie detectors. Because theory specifies that liars behave differently than honest senders, poor detection should be limited to people unaware of which behaviors are diagnostic. People who are naturally adept, who have undergone extensive training, or who are professionally experienced at recognizing and interpreting behavioral signals of deception should—according to the logic of extant theory—be skilled at distinguishing truths from lies. Such individuals might be called lie detection experts.

Examination of the literature on expertise in deception detection reveals four rival perspectives on deception-detection expertise. These include the finding that expertise makes little difference; the idea that experts are inferior because they are lie-biased; the “experts-are-superior-with-high-stakes-lies” argument; and the “detection wizard” concept. Each of these views is discussed in turn. Then, a fifth alternative view is described. We propose that lie detection expertise is a function of actively prompting diagnostic answers to strategic questioning, rather than the passive observation of nonverbal behavior.

Expertise has little impact on detection accuracy

Based purely on past empirical findings, the most well documented and scientifically defensible prediction is that experts and nonexperts differ little in deception-detection accuracy. The Bond and DePaulo (2006) meta-analysis identified 42 deception-detection experiments involving expert judges. The weighted average accuracy for experts was 53.8% compared to 53.3% for nonexperts. Bond and DePaulo also identified 19 studies which compared experts and nonexperts within the same study. In head-to-head comparisons, experts were slightly less accurate than nonexperts ($d = -.03$).

Similar findings of little difference were reported by Aamodt and Custer (2006). Across $k = 156$ studies involving students as judges, accuracy was 54.2% compared to 50.8% for detectives ($k = 7$), 54.5% for federal officers ($k = 4$), and 55.3% for police ($k = 12$). Age correlated with accuracy at $r = -.03$ and years of professional experience correlated with accuracy at $r = -.08$.

Finally, Bond and DePaulo (2008) report that variability in judge accuracy is minimal, and there is no evidence of individual differences in judge ability that might potentially be attributable to expertise or differential judge skill beyond chance variation. Together, these three meta-analyses suggest no consistent evidence that experts are much different from other individuals including students. Instead, the slightly better-than-chance accuracy finding appears robust across age, professional experience, and occupation.

Experts are lie-biased

Although the preponderance of research suggests little or no effect for expertise on detection accuracy, some researchers hypothesize that expertise reduces accuracy. Burgoon, Buller, Ebesu, and Rockwell (1994) were the first to propose the experts-are-inferior view. They reasoned that people with careers involving lie detection develop chronic suspicion. Chronic suspicion makes experts lie-biased, and lie-bias was thought to impair accuracy. Meissner and Kassin (2002) subsequently called the tendency for experts to exhibit lie-bias “investigator bias.”

In Burgoon et al.’s (1994) findings, experts (instructors in a military intelligence school) were less likely to judge senders as honest and were less accurate than nonexpert adults. In Meissner and Kassin’s findings, experts were no different from students in detection accuracy, but experts were more confident in their judgments and were more likely to judge senders as guilty liars. Bond (2007) also reports little difference in accuracy and that experts were more lie-biased than students. Consistent with this view, Masip, Alonso, Garrido, and Anton (2005) reported that experienced police were higher in trait suspicion than college students or new police recruits.

The argument that experts are less accurate because they are lie-biased is implausible for at least two reasons. First, the preponderance of evidence suggests that experts are *not* lie-biased. The Bond and DePaulo (2006) meta-analysis reports that experts are, on average, slightly truth-biased; judging 52.3% of messages as honest. Second, even if the premise of experts being lie-biased was true, the conclusion of

lower accuracy resulting from lie-bias does not logically follow. As Levine, Park, and McCornack (1999) argued in the explication of the “veracity effect,” accuracy in the literature is calculated by averaging across an equal number of truths and lies. If experts were lie-biased, lie bias would reduce accuracy for truthful statements but increase accuracy for lies with the truth-detection impairment and lie-detection gain canceling out. The net result is that neither truth-bias nor lie-bias substantially affects overall accuracy; one type of error is simply traded for another.

Expertise is moderated by lie stakes

A contrasting view is advanced by O’Sullivan, Frank, Hurley, and Tiwana (2009). Whereas Burgoon et al. (1994) argue that experts are inferior, O’Sullivan et al. (2009) claim that at least some experts are adept at detecting deception under conditions of consequential, high-stakes lies. That is, lie stakes and expertise statistically interact, such that high accuracy is observed for experts judging high-stakes lies. When either student samples are used, or when experts judge low-stakes lies, accuracy is just slightly above chance.

O’Sullivan et al. (2009) acknowledge the existence of meta-analyses suggesting poor accuracy and a lack of effects for expertise, but dismiss those findings. As they argue:

This [accuracy is slightly-better-than-chance as reported in meta-analysis] mistaken conclusion may have resulted from the disproportionate representation of college students in most lie detection research. Given that college-age students are significantly less accurate than adults older than 22 [self-citation to an unpublished poster is provided], when college students contribute the bulk of the data examined in a meta-analysis, their chance scores will swamp the statically significant results reported for police groups in several countries. (p. 530)

In support of their conclusions, O’Sullivan et al. reviewed 23 studies examining 31 police groups in deception-detection tasks. Each accuracy task was coded as involving high- or low-stakes lies. They report 11 accuracy results at 60% or above. Ten of the 11 findings above 60% accuracy were coded as high-stakes lies. Of the 16 results at or below the meta-analysis average of 54%, all were coded as low-stakes lies. They report that the (unweighted) mean accuracy for experts judging high-stakes lies was 67.2%, compared to 55.2% accuracy for low-stakes lies. Taken at face value, O’Sullivan et al.’s literature review appears to offer strong and compelling support for the expertise-by-stakes interaction argument.

O’Sullivan et al.’s (2009) portrayal of the literature conflicts sharply with the results from formal meta-analyses, however. Whereas O’Sullivan et al. assert as fact that college students are inferior to older adults over 22, Aamodt and Custer (2006) explicitly tested age as a moderator, and report a nonsignificant, wrong-direction correlation of $r = -.03$ between age and accuracy (upper bound of 95% CI, $r \leq +.01$, 100% of standard error explained by sampling error). Bond and DePaulo (2006) included both expertise and stakes as moderators in their analysis, and neither were substantial moderators. Accuracy for no motivation lies was 53.4% compared to 53.3% for motivated

lies. Further, there was not sufficient heterogeneity in either meta-analysis to account for the two-way interaction suggested by O'Sullivan et al.

Two factors likely explain the discrepancy between the results of large-scale meta-analyses and O'Sullivan et al.'s (2009) smaller and less formal literature review. The first issue is the sample sizes associated with the evidence referenced by O'Sullivan et al. in support of their argument. The three most supportive findings identified by O'Sullivan et al. derive from samples with $N < 24$. All 10 supportive findings of accuracy $> 60\%$ come from $N < 100$ studies. Because estimates based on smaller samples are less stable, sampling error may, in part, explain the departure of the across-study averages. Second and more importantly, O'Sullivan et al. openly and intentionally cherry-picked findings supporting their view. As they wrote in their method: "Where more than one mean accuracy was available for several different tests from the same group of subjects, the lie scenario resulting in the highest mean accuracy was used" (2009, p. 532). For example, they report the accuracy obtained by Porter, Woodworth, and Birt (2000), which was coded as high-stakes, as 77%. The parole officers in the Porter et al. study were tested four times. They performed as poorly as 40.4% and as high as 76.7%, with the average across tests being 60.5%. Had the most conservative value been picked, that data point would have been wrong direction inconsistent with the O'Sullivan et al. argument. Because O'Sullivan et al.'s conclusions rest on the optimistic and opportunistic selection of small sample findings, skepticism of their claims may be prudent.

Wizards

O'Sullivan and Ekman (2004) contend that a very small proportion of experts are exceptionally adept at distinguishing truths from lies at rates better than 80% or 90% across three sets of videotaped truths and lies. In an independent test of the Wizards idea, Bond (2007) tested 112 experts and 122 students on a lie detection task. Overall, students and experts did not differ from each other or from the slightly better than chance accuracy levels reported by meta-analyses. Eleven (10%) of the experts, however, obtained levels of accuracy over 80%. When tested again, 2 of the 11 experts again performed at better than 90%. Although most well-performing experts regress to the mean, some small number of experts exhibit strings of successful judgments that appear statistically improbable.

Expertise and the ecology of deception detection

An alternative explanation for near chance accuracy findings in the literature is ecological. In most deception-detection experiments, all judges have to go on is sender demeanor. The typical research design makes sense from the perspectives of the deception theories discussed previously. Each of those theories posits that deception is detectable based on demeanor. However, those theories have may blinded researchers to other possibilities, and that research designs influenced by those theories do not capture critical elements of the lie detection ecology experienced by

lie detection experts in their professional roles. If the current reasoning is correct, a more ecological research environment would allow experts' skills to manifest in a way that produces notably higher levels of accuracy.

Park, Levine, McCornack, Morrison, and Ferrara (2002) speculated that one reason that people do poorly within lie detection studies is because experimental methods do not reflect how people really detect lies outside the lab. Park et al. asked participants to recall lies that they had detected. Few of the recalled lies were detected based only on sender demeanor. Instead, most lies were detected either by comparing what was said to known truth (either by prior knowledge, by obtaining physical evidence, or by information from third parties) or by getting the liar to confess their deception. These methods are not available in the typical lie detection experiment. Park et al. concluded that the typical deception-detection experiment involves a task dissimilar from deception detection outside of the lab; consequently, the literature may not adequately reflect either the processes or the outcomes of nonresearch lie detection. According to Park et al., lie detection experiments often preclude what people actually do to catch liars.

Experts most certainly use the strategies reported in the Park et al. (2002) findings, and experts may use those strategies more adeptly than nonexperts. A police detective interviewing a suspect, for example, would surely have more to go on than just demeanor. The very fact that a person is a suspect suggests that the police have some evidence to justify their suspicions. Police compare what the person says to what is known about the crime. They do not merely passively watch demeanor, but instead actively and strategically question a suspect. They not only seek informative nonverbal cues, but information that can be compared to known facts. They are not bound to a fixed script; and they can follow up on suspicious answers. They can check alibis and other factual statements. Finally, their goal is not to detect deception *per se* but instead, to get at the truth. They try to get exculpatory information from the innocent, and admissions from the guilty. In short, what makes an expert adept at lie detection is not the passive observation of sender demeanor, but the active prompting of diagnostic information (Levine, Blair, & Clare, 2014).

The task described above is notably different in a host of ways from the task that experts face in most deception-detection experiments. Usually, content and evidence have no utility in the lab (Blair, Levine, & Shaw, 2010). Motive is of little consideration (Levine, Kim, & Blair, 2010). The lies are often about topics different from the lies that experts face in their professions; or are based on mock crimes where the mock perpetrators are different in many important ways from actual criminals. The experts most often passively watch videotaped potential liars. If the experts are allowed to interact with the people being judged, the questions are typically scripted by the researchers. Experts often do not construct the questions, and they cannot go off script to follow up on suspicious answers. If the senders confess, the data are discarded, and not counted as successful detection.

Recent deception-detection research finds that detecting lies based on demeanor can be very misleading, especially to experts (Levine, Blair, & Clare, 2014). Research

finds too that when information relevant to motive is present, it is used by experts and nonexperts alike (Levine, Kim, et al., 2010). Strategic questioning of a sender can aid detection (Levine, Shaw, & Shulman, 2010; Levine, Blair, & Clare, 2014), and the use of content in context (Blair et al., 2010) and the strategic use of evidence (Hartwig, Granhag, Stromwall, & Kronkvist, 2006; Hartwig, Granhag, Stromwall, & Vrij, 2005a) can produce accuracy rates over 70%. What all this recent research shows, we argue, is that the ecology of the lie detection context matters greatly. The typical deception-detection experiment differs in critical ways from what experts do outside the lab and from the type of situation that would make an expert's expertise an asset.

Consistent with the current arguments, research on expertise in tasks other than deception detection shows that expertise is highly bounded by context and domain (Chi, 2006). Chess masters, for example, demonstrate a remarkable ability to remember positions that actually occur in chess games; but are much less accurate at reproducing randomly generated chess positions (Gobet & Simon, 1996). Feltoich and Barrows (1984) found that physicians who are given background information about a patient that was not necessarily causal to the current condition were 50% more accurate in making a diagnosis than physicians that were simply given pictures of the patient, patient charts, and a list of complaints. The extensive experience of experts appears to allow them to use a wide array of correlational knowledge to improve their judgments in much the same way that Brunswick (1955) argued that people make use of many fallible indicators to arrive at judgments which far exceed the accuracy possible from the use of a single indicator. In short, the research on expertise suggests that deception-detection expertise would be both limited to a specific domain and require access to the contextual clues that normally exist within that domain.

Research predictions

We propose that expertise in deception detection manifests as the context dependent ability to actively prompt diagnostic answers rather than the passive ability to appraise the meaning of observed sender demeanor. In this view, neither mere interaction nor mere question asking is sufficient to enhance accuracy. The literature clearly suggests that mere interactivity is insufficient to enhance accuracy. Buller, Strzyzewski, and Hunsaker (1991), for example, found that passive observers were 59% accurate compared to 55% accuracy for conversational participants. A similar nonsignificant trend was reported by Burgoon, Buller, and Floyd (2001).

This pattern was reported in the Bond and DePaulo (2006) meta-analysis, where studies in which the judge was a passive observer had a weighted mean accuracy of 54.0% compared to 52.8% accuracy in studies in which the judge interacted with the sender; and the lack of difference extends to police as questioners and observers (Hartwig, Granhag, Stromwall, & Vrij, 2005b). If anything, judges interacting with senders are slightly less accurate than passive observer judges. Nevertheless, the slightly better-than-chance rate applies to both passive observers and judges engaged in active interaction.

Similarly, merely asking a potentially deceptive message source questions has little impact on accuracy. In a finding labeled the “probing effect,” research has repeatedly found that merely asking probing questions of a sender does not affect detection accuracy and instead makes the sender more believable (Levine & McCornack, 2001). Expertise is required to ask questions in a way that will yield answers that are informative about the truth.

We hypothesize that using a research design that is more similar to the situations facing experts will yield accuracy rates significantly and substantially higher than that reported by meta-analyses. Such an ecological design requires a number of features. First, in contrast to Hartwig et al. (2005b), the potential lies being judged need to involve covering up an actual misdeed (rather than role-playing a mock crime) and should not be sanctioned by the research design. Second, obtaining confessions needs to be a viable approach to uncovering the truth. Third, the expert needs familiarity with the context. Fourth, the expert constructs a question script, but has the liberty to stray from the script. Finally, the expert interacts directly with the suspects. We anticipate that under such conditions, a well trained and experienced expert will do substantially better than chance at distinguishing guilty from innocent suspects.

A second hypothesis is that nonexpert passive observers will be well above chance accuracy when judging expertly questioned senders. The reasoning behind this prediction is based on the premise that deception-detection expertise manifests not as passive observation but instead as active elicitation of diagnostic answers. Thus, if the expert prompts answers that make the truth more apparent, the benefits should apply not only to the expert, but also to passive nonexpert observers.

These predictions differ from the four perspectives reviewed previously. We anticipate accuracy well above the levels reported in meta-analyses. Unlike Burgoon et al. (1994), we do not expect the expert to be either lie-biased or inferior to nonexperts. Counter to O’Sullivan et al. (2009) and the wizards approach, we do not expect the improved accuracy to be limited to our expert. Instead, we expect that a student sample viewing expertly questioned senders can achieve elevated levels of accuracy beyond that suggested by meta-analyses.

Method

Two studies were conducted, each involving three phases. The first study provided proof-of-concept that was used to obtain additional funding for the second study. The second study replicated the initial findings with different experts, different interviewees, and different passive student judges, thus providing a full cross-validation.

In the first phase of each study, subjects were given the opportunity to cheat on a task for a cash prize. The decision to cheat or not was each subject’s own to make. The research design simply provided opportunity and incentive for a transgression. Because the cheating was done by students at a university, because the subjects believed that the cheating would harm the results of the research, and because the cheating involved the attempted theft of federal research funds, cheaters had

motivation to cover up their transgression and noncheaters were motivated not to be seen as cheaters. Although cheaters were not actually punished (beyond the loss of ill-gotten funds), cheaters were unaware of the lack of consequences until the debriefing that took place immediately following the completion of their phase two interview.

During the second phase, subjects who had just completed the first phase were interviewed about Phase 1. Unknown to the subjects, the interviewer was a trained and experienced expert interrogator. The expert was blind to whether the subject actually had cheated, and attempted to ascertain (via questioning) if cheating had occurred. Following each interview (which averaged approximately 4 minutes in duration in the first experiment and 9 minutes in the second study), the expert recorded his or her opinion about the innocence or guilt of each subject. All interviews were videotaped for use in Phase 3.

The third phase involved showing the videotaped interviews from Phase 2 to samples of undergraduate student judges who were asked to judge the guilt or innocence of each interviewee. The judgments were scored for accuracy and compared to previous findings from meta-analyses. All phases of the research were IRB approved, and all members of the research team were human subjects certified.

Participants

The participants in Phase 1 of Study 1 were $N = 33$ (10 females, 23 males) undergraduate students enrolled in a large multisection freshman-level service class at a large university in the midwestern United States. Class research credit and the chance to win up to \$50 USD were offered in exchange for less than 1 hour of participation. Phase 1, Study 2 participants were $N = 89$ (46 females, 43 males) students recruited from the same class 2 years later. The incentive for participation was increased to a chance to win up to \$100.

Study 1, Phase 3 participants were $N = 136$ (56.6% female, mean age = 19.2, $SD = 0.13$) undergraduate students enrolled in communication classes at the same university. Class research credits were offered in exchange for participation. Study 2 involved an additional $N = 34$ student judges (71.4% female, mean age = 21.32). No participant from Phase 1 participated in Phase 3.

The trivia game

A trivia game was used to provide participants with the opportunity and incentive to transgress by cheating. The trivia game has been successfully used to prompt high-stakes lies in several published studies (e.g., Blair et al., 2010; Levine, Kim, Park, & Hughes, 2006; Levine, Kim, et al., 2010; Levine, Shaw, et al., 2010) and was loosely modeled on a procedure originally created by Exline, Thibaut, Hickey, and Gumpert (1970). Participants played a trivia game with a partner for a cash prize of \$5 each for each correct answer (\$10 each in Study 2). Stacks of \$5 or \$10 dollar bills were placed in front of participants and partners, and a bill was removed after each missed

question. The questions were difficult. The partner was a trained research confederate acting as another subject.

Between the third and fourth questions, the experimenter was called out of the room, leaving the money and the answers (in a folder) behind. In Study 1, the participant's partner suggested cheating, but did not excessively pressure the participant to cheat. Partners only cheated if participants both agreed to cheat and actively participated in the cheating. In Study 2, there were three randomly assigned conditions. Either the confederate did not mention cheating and only cheated if cheating was instigated by the true subject, the confederate cheated regardless of the subject's wishes and asked the subject not to snitch on them, or the confederate tried to solicit partner cheating as in Study 1. Upon the experimenter's return, the game resumed and was eventually completed when all 10 questions had been asked. Following the trivia game, participants engaged in a videotaped interview in an adjacent room.

The experts

In Study 1, a male expert questioned each of the 33 participants completing Phase 1. The expert's qualifications included being certified in the Reid Technique, being a former Reid Technique instructor for John E. Reid and Associates, and being a former professional investigator for John E. Reid and Associates. The expert holds a Ph.D. in criminal justice and has read and published original research on deception detection, interrogation, and false confessions. The expert has trained federal, state, and local law enforcement and military personnel in interview and investigation techniques. The expert was familiar with research on question effects, content in context, and the strategic use of evidence; as well as previous iterations of the cheating study.

The experts in Study 2 were five federal agents with substantial polygraph and interrogation experience. Each participated voluntarily in exchange for a small honorarium and having their travel expenses paid. Four of the agents were male and one was female. All experts were blind to whether participants actually cheated; but they were thoroughly briefed on the experimental setting and cheating game setup.

The interviews

The interviews in Study 1 were modeled after an abbreviated version of the Behavioral Analysis Interview (Inbau, Reid, Buckley, & Jayne, 2001; Horvath, Blair, & Buckley, 2008), and lasted, on average, about 4 minutes. The interviewer began by asking the student's name, major, their year in school, and where they were from. This personal identification information was subsequently edited out of the interviews prior to Phase 3. The participant was then asked to describe in detail what had happened during Phase 1 of the experiment. They were asked how they did on the trivia questions, which questions they got right, and—for the questions they knew—*how* the answers were known. The participant was then asked if he or she had cheated. If the participant denied cheating, s/he was asked what should happen to someone who did

cheat and if there were circumstances in which s/he would give a cheater a break. Some participants were then asked if someone was going to cheat, how they would go about it. Finally, a bait question was asked (Inbau et al., 2001). The participant was informed that his or her partner would be interviewed next, and asked what the partner was going to say. Following that question, the experimenter excused himself for a few minutes to “check on some things.”

Upon return, the participant was told that it was clear that he or she had cheated. The interviewer told the participant that he wanted to know why s/he had cheated. It was explained that it was one thing if s/he had done it simply to win some money, but another thing if s/he had been intentionally trying to sabotage the experiment. It was explained that the important thing was the integrity of data, and that was all the interviewer was interested in. Otherwise, there was going to be an investigation. The participant was then given the opportunity to be “up front” with the interviewer and confess that she or he had done it for the money. If the participant still denied cheating, s/he was told that his or her partner had said s/he had cheated. The interview ended either with a confession or continued denial. Immediately following the interview, the expert recorded his opinion about the participant’s guilt or innocence.

The interviews in the second experiment were completely unscripted and varied substantially in style and substance from expert to expert. The interviews ranged in duration from 2 minutes, 57 seconds to 17 minutes, 44 seconds ($M = 9:05$, $SD = 3:39$). All participants were immediately debriefed following the interview and no adverse events occurred.

Phase 3

In Study 1, all the interviews in Phase 2 were videotaped and shown to new participants in Phase 3. All personal identifying verbal content was edited out of each interview, as was the time while the interviewer was out of the room. The tape provided a full body and head view of the seated interviewee with the expert off camera. Because there were 33 tapes which averaged about 4 minutes each, the interviews were divided into five approximately half-hour-long compilations, each containing five to seven interviews. Participants watched one of the five compilation tapes judging each interviewee as guilty or innocent of cheating. Guilty and innocent judgments were scored for percent accuracy and percent judged innocent. The particular compilation viewed did not impact accuracy, $F(4, 131) = 0.95$, $p = .44$ and the results summed across compilations.

In Study 2, $N = 34$ students enrolled in an undergraduate special topics class on deception watched and rated a random sample of 36 (17 in which cheating occurred, 19 no cheating occurred; 17 males and 19 females) of the Study 2 interviews as a class project.

Proof-of-concept results

In Phase 1 of Study 1, 4 of the 33 participants (12%) cheated. The rate of cheating was about half of that obtained in a previous version of the cheating induction (24%,

Levine, Kim, & Hamel, 2010). All four cheaters (100%) confessed under expert questioning; compared with a 40% confession rate reported by Levine, Kim, and Hamel (2010). Two (50%) of the confessions were elicited with the false evidence ploy. A third confession occurred earlier in the interview, when the participant was explicitly asked if he or she had cheated. The remaining confession occurred even earlier as part of the description of what happened in the experiment. As in previous versions of the cheating induction (e.g., Levine, Kim, & Hamel, 2010), no innocent participant confessed. That is, the false confession rate was 0.0%.

The expert interviewer's judgments were tallied and the expert was correct on all 33 (100%) interviewees. A binomial test was used to test the hypotheses that the experts hit rate was greater than 50/50 chance, and the 54% meta-analyses level. The probability of 33 hits on 33 successive trials presuming a .5 hit rate is approximately a hundred-million-to-one and the probability given a .54 hit rate is approximately one-in-ten-million. The probability of judging 29 of 33 (88%) senders as honest presuming the expert was lie-biased was less than three-in-ten-thousand. The probability of correctly identifying all 33 participants assuming that the expert was able to incorporate base-rate information from previous runs of the cheating experiment (.76 truthful) was approximately one-in-ten-thousand. Thus, the expert was statistically better than 50/50 chance, the 54% meta-analyses level, and the base rate in previous cheating experiments; and the lie-bias hypothesis was statistically improbable.

The Phase 3 participants judged 65.5% ($SD = 15.5$, 95% CI = 62.6 to 68.4) of senders as innocent and obtained 79.1% average accuracy ($SD = 16.5$, 95% CI = 76.0 to 82.2). One sample t tests showed that the observed level of accuracy was statistically greater than 50–50 chance, the .54 meta-analysis mean, and the 0.655 honest guess rate; $t(135) = 13.7$, $p < .0001$, $t(135) = 11.1$, $p < .0001$, and $t(135) = 7.07$, $p < .001$. The modal accuracy was 100% with 23.5% of participants obtaining perfect accuracy. The median was 83.3%. Accuracy on innocent senders was 76.7% and accuracy on guilty cheaters was 92.7%. Three of the confessors were almost universally judged guilty (100%, 100%, and 97%) while the fourth was judged guilty 76% of the time.

Cross-validation results

In Phase 1 of the second study, cheating occurred on 40 of the 89 trivia game sessions (44.9%). In all, no cheating occurred on 49 instances (55.1%), the confederate but not the participant cheated on 13 occasions (14.6%) and both the participant and the confederate cheated on 27 trials (30.3%). Confessions were successfully solicited for 34 of 40 (85%) sessions in which cheating occurred, with six participants maintaining false denials for the duration of the interview. As in previous versions of the cheating induction (e.g., Levine, Kim, & Hamel, 2010), no innocent participant confessed. That is, the false confession rate was 0.0%.

The expert interviewers' judgments regarding if any cheating occurred were tallied and the experts were correct on 87 of 89 interviews ($M = 97.7\%$, $SD = 14.2\%$, 95%

CIs = 94.5%–100.9%). Of the two errors, one was a false positive and the other was a false negative. Accuracy levels for the individual interviewers ranged from 94.7% to 100% with 3 experts obtaining perfect accuracy over 14, 16, and 20 interviews each. A binomial test was used to test the hypotheses that the experts' rate of correctly distinguishing cheating from noncheating sessions was greater than 50/50 chance and the 54% meta-analyses level. The probability of 87 hits on 89 successive trials presuming a .5 hit-rate is $6.3E^{-24}$ and the probability given a .54 hit-rate is approximately $4.3E^{-21}$. Accuracy was not correlated with the duration of the interview ($r = -.05, p = ns$). The experts judged 55.1% of the interviewees as not involving cheating either by the subject or by the confederate. This matched the actual rate of noncheating (55.1%). Thus, the experts were statistically much better than 50/50 chance and the 54% meta-analyses level; and no evidence for the lie-bias hypothesis was observed.

Experts were also assessed on their ability to correctly identify who cheated. There were three possibilities. Confederates may have cheated without participant involvement; both the confederate and the participant might have cheated; or neither of them cheated. The experts correctly identified the actual culprit or culprits on 85 of 89 (95.5%) interviews. In addition to the two previously mentioned errors regarding if cheating occurred at all, the two additional errors involved experts erroneously concluding that only the confederate had cheated when, in fact, both the confederate and the participant cheated.

The Phase 3 participants judged 55.1% ($SD = 6.8, 95\% CI = 52.9-57.3$) of senders as innocent and obtained 93.6% average accuracy ($SD = 6.8, 95\% CI = 91.4-95.8$). One sample t tests showed that the observed level of accuracy was statistically greater than 50–50 chance, $t(34) = 38.3, p < .0001$, and the .54 meta-analyses mean, $t(34) = 34.8, p < .0001$. The modal accuracy was 100% with 27.8% of participants obtaining perfect accuracy. The median was 95.0%. Accuracy on senders where cheating did not occur was 93.4% ($SD = 7.8\%$) and accuracy on correctly identifying if cheating occurred was 94.1% ($SD = 7.8$).

Discussion

This research involved deception-detection experts actively interviewing potential cheaters with the intent of accurately assessing guilt or innocence based on the expert's interaction with the potential cheaters. The Study 1 expert obtained perfect 100% accuracy over a string of 33 successive interviews. Videotapes of the interviews were shown to student judges who obtained, on average, 79% accuracy. Almost one quarter (23.5%) of the student judges obtained 100% accuracy and half the student judges scored 83% or better. The findings were subsequently replicated with new experts, interviewees, and student judges in the second experiment. Five experts obtained 98% accuracy over 89 interviews, and student judges watching a sample of the interviews also obtained 94% accuracy.

The current results are sharply inconsistent with the results of prior deception-detection research. Meta-analyses clearly document previous findings indicating that (a) deception detection is only slightly better than 50–50 chance, and (b) neither experts nor nonexperts are adept lie detectors. Of course, the current results do not invalidate the meta-analytic findings. The previous results are very well documented and robust within the research designs that produced those results. Instead, the current findings clearly come from a statistically different population of judgments than the findings that populate the existing deception-detection meta-analyses.

The critical constellation of differences, we believe, is related to the *ecology of the deception-detection task*. In past research, experts were typically passive observers with little to go on other than sender demeanor. In the current studies, the experts were active and adaptive agents in a familiar context. The experts were able to ask questions that promoted diagnostic responses. The critical aspect of expertise was not in the expert's ability to read body language or microfacial expressions; but instead, was knowing what questions to ask and how and when to ask those questions. The responses to the experts' questions were sufficiently diagnostic that even nonexperts passively watching the interviews on videotape were able to distinguish guilt from innocence at rates well above meta-analytic levels. In fact, the lower bound of the confidence interval for student judges in the first study (76%) was higher than the highest accuracy rate (73%) from any individual study reported in the Bond and DePaulo (2006) meta-analysis. In looking at the distribution of student judges accuracy scores, 23.5% and 27.8% of the students obtained 100% perfect accuracy. In Study 1, less than 10% of student judges performed at or below the mean meta-analytic level of 54% accuracy. Student judge accuracy was even higher in the second study (95% CIs = 91.4–95.8).

The current results were inconsistent with the experts are lie-biased (or investigator biased) hypothesis. The expert investigator in Study 1 made no false positive errors. In Study 2, false positives and false negatives were equal. No false confessions were elicited in either study. In Study 1, because an unusually high percentage of participants in Phase 1 were innocent and honest about their innocence, if our expert was guilty or lie-biased, his accuracy would have been low. Of course, because our six experts were not lie-biased does not mean that all trained and experienced experts are similarly unbiased. But, our experts were sensitive to the actual guilt-innocence base rate even though the guilt rate varied. Further, meta-analyses are inconsistent with the guilt and lie-bias prediction. Finally, lie-bias has not been observed in previous studies with experts judging the cheating tapes (e.g., Blair et al., 2010; Levine et al., 2014; Levine, Kim, et al. 2010). Therefore, the conclusion that experts are guilty or lie-biased appears to be an overgeneralization lacking robust empirical documentation. We do not doubt that experts are sometimes guilty or lie-biased, but we do question the pervasiveness of bias.

The performance of our experts was in line with both the stakes-by-expertise perspective and the lie detection wizards perspective. The lies in the current research were high-stakes, so the high performance by the experts is consistent with the

stakes-as-a-moderator argument. It is also possible (although statistically improbable) that our experts were all wizards. What is inconsistent with these views, however, is that the two student samples performed so well. That 23.5% and 27.8% of the college students obtained perfect 100% accuracy shows that once diagnostically useful answers were elicited by the expert, neither expertise nor wizardly ability was required for high accuracy. This suggests that expertise rests in the ability to elicit diagnostic answers, rather than the ability to divine the meaning behind the answers or the behaviors accompanying the answers.

Perhaps the most intriguing question is what made the experts so good? We believe there were two facets of the task that were critical in the experts obtaining high accuracy. First, the experts were highly adept at obtaining true confessions while not soliciting false confessions. In the current studies, confessions were highly diagnostic. Judging confessions as an indication of cheating and consistent denials as indicative of noncheating would yield 100% accuracy in Study 1 and 93% accuracy in Study 2.

The second element was the use of what Blair et al. (2010) call “content in context.” In the cheating studies, the number of correct judgments is highly diagnostic. The questions were difficult, and a simple decision rule that more than 1 correct answer indicates cheating yields an accuracy level of approximately 80%. Only nine noncheating teams obtained more than 1 right answer and all but 6 pairs of cheaters obtained two or more correct answers. It was clear from the lines of questioning that the experts were interested in the number of questions answered correctly, who got the questions right (the participant or the confederate), and how the participant happened to know the right answer.

Besides the findings of high accuracy, three additional findings are particularly noteworthy and newsworthy. First, the variability in accuracy observed in Phase 3 of both studies was markedly different from the pattern typically observed in the literature, and this difference is theoretically and practically informative. In the typical deception-detection experiment, the variance in sender transparency is larger by a factor of 10 than variance in judge ability (Bond & DePaulo, 2008). In the current results, that difference was not observed. The difference in findings can be seen by comparing the current findings to a previous experiment using the cheating induction with a nonexpert interviewer. Consistent with the Bond and DePaulo (2008) meta-analysis, Levine, Shaw, et al. (2010) reported that variance for senders (615) was much larger than for judges (44). In the current results, the variance in senders was substantially reduced (228) and the variance for judges was notably increased (272).

The larger variability for senders in the previous work was reflective of the fact that there are usually some senders that almost all judges got wrong, and others in which almost all judges scored correctly (see Levine et al., 2011 for an explication and documentation of these individual differences in sender demeanor that are independent from actual veracity). The small variance in judges reflects the lack of effects for judge ability that is typical of the literature. All judges usually perform similarly.

The responses elicited by the expert in the current study made most senders much more transparent than usual. No sender in the current data yielded across-judge accuracy below 46% and almost half the senders were judged correctly 80% of the time or more. This reduced the variance in senders relative to previous detection studies.

At the same time, the prompting of diagnostic information instilled variance in judge accuracy. Because sender responses were diagnostic, judge ability and motivation mattered. This enabled some judges to do better than others who were less skilled or less motivated. Thus, the expert questioning reduced the number of senders who were difficult to detect and made judge ability and motivation more important.

A second provocative finding was that no false confessions were observed. The lack of false confessions is consistent with previous iterations of the cheating induction. In previous experience with interviewing potential cheaters, true confessions were reasonably common. However, across approximately 400 trials over 6 years, not a single false confession has happened as a consequence of the various questioning strategies implemented. This lack of false confessions is noteworthy, because minimization, maximization, and false evidence strategies were used by the Study 1 expert in the current questioning. In past research, strategies such as these have produced false confessions in a conceptually similar research design (e.g., Russano, Meissner, Narchet, & Kassin, 2005). Obviously, some moderator(s) are at play.

One possible difference is the degree of perceived transgression. In Russano et al. (2005), cheaters were helping another research subject, while in the current study cheaters were motivated by self-gain. Further, talking to another subject in defiance of instructions may have seemed a less serious offense than looking through obviously private materials in order to win money. Perhaps the perceived severity of the offense made false confessions less likely as was found by Horselenberg, Merckelbach, and Josephs (2003).

A more likely moderator may be experimenter demand effects. The purpose of the Russano et al. (2005) and other false confession studies are to obtain false confessions. If no false confessions are obtained, the study is a failure. The purpose of the current study was to enhance accuracy. False confessions would have worked against the intent of the current study. The experts in the current study were trying to produce true confessions and avoid false confessions. If the current goal had been to obtain false confessions rather than to get at the truth, we suspect that at least some false confessions could have been obtained.

This suggests that interviewer intent may be the critical factor in the production of false confessions. An interrogator intent on obtaining a confession regardless of the truth may get false confessions, while an interviewer intent on discovering the truth may be less likely to generate false confessions even if they use minimization and false evidence strategies. That is, false confessions may be less a function of the strategy used to gain the confession *per se* than the intent behind the strategy.

A third observation that merits discussion is the relatively short duration of the study one interviews. The duration of the interviews was only 4 minutes, yet that was sufficient time to prompt confessions from guilty cheaters and to allow innocent

participants to communicate their innocence effectively. The expert questioning was not only efficacious, it was also efficient. A novice or a fully scripted interviewer would be unlikely to perform so well in so little time, and the relatively short duration of the interviews makes the findings all the more impressive.

The current findings have important theoretical implications. Most deception theories specify that recognizable nonverbal cues to deception stem from the cognitive and affective consequences of telling a consequential lie, and those cues should be recognizable to the savvy passive observer. The current findings suggest a different view in which cues do not leak on their own, but can be prompted by someone skilled in asking strategic questions. Further, the nature of diagnostic cues is not a matter of correctly reading sender demeanor, but instead of listening to communication content in context (Blair et al., 2010). Consistent with truth-default theory (Levine, 2014), lies are detected and innocence is ascertained by comparing what is said to what is known and by the elicitation of confessions. This is how people detect lies outside the lab (Park et al., 2002), and this is just what the experts expertly did in the current studies.

The current research can be compared and contrasted with Vrij and Granhag's (2012) commentary on the state of deception-detection research in legal psychology. Both approaches advocate for more ecological approaches to research design, both see value on researcher-practitioner collaboration, both highlight the importance of asking the right questions, and both focus on actively prompting diagnostic information. There are also several noteworthy differences. The current research is focused on bottom-line accuracy while Vrij and Granhag (2012) focus on cue elicitation. The mechanisms in Vrij and Granhag involve cognitive load and the strategic use of evidence while the current research explains improved accuracy by attention to communication content and solicitation of honesty. Vrij and Granhag see little value in students watching video tapes. We disagree, and believe our current student data contribute meaningfully to our results. Although our theoretical approach differs from Vrij and Granhag and Dunbar et al. (2013), we are certainly not alone in the quest for improved deception detection and ecological research design.

An obvious limitation of the current experiment is the lack of maintained lies. In fact, in Study 1, given that no sender was able to maintain his or her lies throughout the entire interview, some might argue that the current research is not about deception detection at all. All the tapes, after all, contained either honest denials or honest confessions. Although more guilty senders overall and some guilty senders who maintained their innocence would have made the studies more comparable to the base rates typical in the literature, a number of considerations mitigate against concerns about the truth-lie base rate and mere truth-bias as explanations for the findings of high accuracy.

From the perspective of investigator bias and expectancy effects, the low number of cheaters likely worked against high accuracy for the expert, making the findings more, not less, compelling. Even if investigator bias view is rejected, however, there is no evidence that experts are strongly truth-biased. The relative infrequency of guilty interviewees and the lack of maintained lies cannot account for the experts' success.

A critic might argue, however, that while the mostly truthful-sender interviews might not account for the experts' accuracy, it might explain student accuracy. Students, are, after all, known to be truth-biased (Bond & DePaulo, 2006; Levine et al., 1999). Previous research with a nonexpert interviewer, however, shows that (a) students in past studies were insensitive to truth-lie base-rates, and (b) merely providing all truthful interviewers was insufficient to produce the levels of accuracy observed in current study (Levine et al., 2006). Additionally, the heuristics and biases research has consistently shown that people either ignore or underutilize base-rate information when making judgments (Kahneman, Slovic, & Tversky, 1982). Base-rate sensitivity is, therefore, not a compelling explanation for the current findings.

Some readers might have concerns regarding the generality of the findings. One might wonder about the generality of the current findings to different contexts involving different misdeeds by different types of wrongdoers. One might wonder how unique the findings are to the individual experts used. One might wonder how many other experts share the level of skill demonstrated by the particular experts in this study. We could speculate about each of these counterfactual possibilities, and we hope that some of these questions will be addressed in additional future replications. Nevertheless, we believe that questioning the generality of the current finding, while natural, misses the main point of the research. Our aim is not make sweeping generalizations, but to test generalizations (cf. Mook, 1983). This research sought to test if meta-analytic findings, claims of investigator lie-bias, and low expert-deception-detection-accuracy would extend to a more ecologically sensitive research environment where the expert was an active agent in a familiar context trying to elicit the truth about potential misdeeds from potential wrongdoers.

In conclusion, the thesis of this paper is captured well by the paper title. It was proposed that expertise in deception detection involves actively promoting diagnostic information rather than passive observation of sender demeanor. This was tested by having experts question more than 100 potential cheaters. The experts were nearly perfect at distinguishing innocence from guilt, and student judges watching videotapes of the interviews were highly proficient as well. These findings suggest that high levels of deception detection may be possible, but require that the right questions are asked the right way in a situation where message content is useful and where the solicitation of honesty is a viable strategy.

References

- Aamodt, M. G., & Custer, H. (2006). Who can best catch a liar. *Forensic Examiner*, *15*, 6–11.
- Bem, D. J. (2011). Feeling the future: Experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology*, *100*, 407–425.
- Blair, J. P., Levine, T. R., & Shaw, A. J. (2010). Content in context improves deception detection accuracy. *Human Communication Research*, *36*, 423–442. doi:10.1111/j.1468-2958.2010.01382.x.
- Bond, C. F., Jr., & DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality and Social Psychology Review*, *10*, 214–234. doi:10.1207/s15327957pspr1003_2.

- Bond, C. F., Jr., & DePaulo, B. M. (2008). Individual differences in judging deception: Accuracy and bias. *Psychological Bulletin*, **134**, 477–492. doi:10.1037/0033-2909.134.4.477.
- Bond, G. D. (2008). Deception detection expertise. *Law and Human Behavior*, **32**, 339–351.
- Brunswick, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review*, **62**, 193–217.
- Buller, D. B., & Burgoon, J. K. (1986). Interpersonal deception theory. *Communication Theory*, **6**, 203–242.
- Buller, D. B., Strzyzewski, K. D., & Hunsaker, F. G. (1991). Interpersonal deception: II. The inferiority of conversational participants as deception detectors. *Communication Monographs*, **58**, 25–40. doi:10.1080/03637759109376212.
- Burgoon, J. K., Buller, D. B., Ebesu, A. S., & Rockwell, P. (1994). Interpersonal deception: V. Accuracy in deception detection. *Communication Monographs*, **61**, 303–325. doi:10.1080/03637759409376340.
- Burgoon, J. K., Buller, D. B., & Floyd, K. (2001). Does participation affect deception success? A test of the interactivity principle. *Human Communication Research*, **27**, 503–534. doi:10.1093/hcr/27.4.503.
- Chi, M. T. H. (2006). Two approaches to the study of experts' characteristics. In K. A. Ericsson, N. Charness, P. J. Feltovich & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and performance* (pp. 21–30). Cambridge, England: Cambridge University Press.
- Dunbar, N. E., Jensen, M. L., Burgoon, J. K., Kelly, K. M., Harrison, B. J. A., & Bernard, D. R. (2013). Effects of veracity, modality, and sanctioning on credibility assessment during mediated and unmediated interviews. *Communication Research*. doi:10.1177/0093650213480175(online first, March 18, 2014).
- Ekman, P., & Friesen, W. V. (1969). Nonverbal leakage and clues to deception. *Psychiatry*, **32**, 88–106.
- Exline, R. V., Thibaut, J., Hickey, C. B., & Gumpert, P. (1970). Visual interaction in relation to Machiavellianism and an unethical act. In R. Christie & F. L. Geis (Eds.), *Studies in Machiavellianism* (pp. 53–76). New York, NY: Academic Press.
- Feltovich, P. J., & Barrows, H. S. (1984). Issues of generality in medical problem solving. In H. G. Schmidt & M. L. de Volder (Eds.), *Tutorials in problem-based learning* (pp. 128–142). Maastricht, Netherlands: Van Gorcum.
- Gobet, F., & Simon, H. A. (1996). Recall of rapidly presented random chess positions is a function of skill. *Psychometric Bulletin and Reviews*, **3**, 159–163.
- Hartwig, M., Granhag, P. A., Stromwall, 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., Stromwall, L. A., & Vrij, A. (2005a). Detecting deception via strategic disclosure of evidence. *Law and Human Behavior*, **29**, 469–484. doi:10.1007/s10979-005-5521-x.
- Hartwig, M., Granhag, P. A., Stromwall, L. A., & Vrij, A. (2005b). Police officers' lie detection accuracy: Interrogating freely versus observing video. *Police Quarterly*, **7**, 429–456.
- Horselenberg, R., Merckelbach, H., & Josephs, S. (2003). Individual differences in false confessions: A conceptual replication of Kassin and Keichel (1996). *Psychology, Crime, and Law*, **9**, 1–8.
- Horvath, F. S., Blair, J. P., & Buckley, J. P. (2008). The behavioural analysis interview: Clarifying the practice, theory and understanding of its use and effectiveness.

- International Journal of Police Science and Management*, **10**, 101–118. doi:10.1350/ijps.2008.10.1.101.
- Inbau, F. E., Reid, J. E., Buckley, J. P., & Jayne, B. P. (2001). *Criminal interrogation and confessions* (4th ed.). Gaithersburg, MD: Aspen.
- Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. New York, NY: Cambridge University Press.
- Levine, T. R. (2010). A few transparent liars: Explaining 54% accuracy in deception detection experiments. *Communication Yearbook*, **34**, 40–61.
- Levine, T. R. (2014). Truth-default theory (TDT): A theory of human deception and deception detection. *Journal of Language and Social Psychology*(online first, May 2014).
- Levine, T. R., Blair, J. P., & Clare, D. D. (2014). Diagnostic utility: Experimental demonstrations and replications of powerful question effects in high stakes deception detection. *Human Communication Research*, **40**, 262–289.
- Levine, T. R., Kim, R. K., & Blair, J. P. (2010). (In)accuracy at detecting true and false confessions and denials: An initial test of a projected motive model of veracity judgments. *Human Communication Research*, **36**, 81–101. doi:10.1111/j.1468-2958.2009.01369.x.
- Levine, T. R., Kim, R. K., & Hamel, L. M. (2010). People lie for a reason: An experimental test of the principle of veracity. *Communication Research Reports*, **27**, 271–285.
- Levine, T. R., Kim, R. K., Park, H. S., & Hughes, M. (2006). Deception detection accuracy is a predictable linear function of message veracity base-rate: A formal test of Park and Levine's probability model. *Communication Monographs*, **73**, 243–260.
- Levine, T. R., & McCornack, S. A. (2001). Behavioral adaptation, confidence, and heuristic-based explanations of the probing effect. *Human Communication Research*, **27**, 471–502. doi:10.1111/j.1468-2958.2001.tb00790.x.
- Levine, T. R., Park, H. S., & McCornack, S. A. (1999). Accuracy in detecting truths and lies: Documenting the "veracity effect." *Communication Monographs*, **66**, 125–144. doi:10.1080/03637759909376468.
- Levine, T. R., Serota, K. B., Shulman, H., Clare, D., Park, H. S., Shaw, A. S., Shim, J. C., & Lee, J. H. (2011). Sender demeanor: Individual differences in sender believability have a powerful impact on deception detection judgments. *Human Communication Research*, **37**, 377–403.
- Levine, T. R., Shaw, A., & Shulman, H. (2010). Increasing deception detection accuracy with strategic questioning. *Human Communication Research*, **36**, 216–231. doi:10.1111/j.1468-2958.2010.01374.x.
- Masip, J., Alonso, H., Garrido, E., & Anton, C. (2005). Generalized communicative suspicion (GCS) among police officers: Accounting for the investigator bias effect. *Journal of Applied Social Psychology*, **35**, 1046–1066. doi:10.1111/j.1559-1816.2005.tb02159.x.
- McCornack, S. A. (1997). The generation of deceptive messages: Laying the groundwork for a viable theory of interpersonal deception. In J. O. Greene (Ed.), *Message production: Advances in communication theory* (pp. 91–126). Mahwah, NJ: Erlbaum.
- McCornack, S. A., Morrison, K., Paik, J. E., Wisner, A., & Zhu, X. (in press). Information Manipulation Theory 2 (IMT2): A propositional theory of deceptive discourse production. *Journal of Language and Social Psychology*.
- Meissner, C. A., & Kassin, S. M. (2002). "He's guilty": Investigator bias in judgments of truths and deception. *Law and Human Behavior*, **26**, 469–480. doi:10.1023/A:1020278620751.

- Mook, D. G. (1983). In defense of external invalidity. *American Psychologist*, **38**(4), 379–387. doi:10.1037/0003-066X.38.4.379.
- O'Sullivan, M., & Ekman, P. (2004). The wizards of deception detection. In P.-A. Granhag & L. A. Stromwall (Eds.), *The detection of deception in forensic contexts* (pp. 269–286). Cambridge, England: Cambridge University Press. doi:10.1017/CBO9780511490071.012.
- O'Sullivan, M., Frank, M. G., Hurley, C. M., & Tiwana, J. (2009). Police lie detection: The effects of lie scenario. *Law and Human Behavior*, **33**, 530–538. doi:10.1007/s10979-008-9166-4.
- Park, H. S., Levine, T. R., McCornack, S. A., Morrison, K., & Ferrara, M. (2002). How people really detect lies. *Communication Monographs*, **69**, 144–157. doi:10.1080/714041710.
- Porter, S., Woodworth, M., & Birt, A. R. (2000). Truth, lies and videotape: An investigation of the ability of federal parole officers to detect deception. *Law and Human Behavior*, **24**, 643–658. doi:10.1023/A:1005500219657.
- Russano, M. B., Meissner, C. A., Narchet, F. M., & Kassin, S. M. (2005). Investigating true and false confessions within a novel experimental paradigm. *Psychological Science*, **16**, 481–486. doi:10.1111/j.0956-7976.205.01560x.
- Vrij, A., & Granhag, P. A. (2012). Eliciting cues to deception and truth: What matters are the questions asked. *Journal of Applied Research in Memory and Cognition*, **1**, 110–117. doi:10.1016/j.jarmac.2012.02.004.
- Vrij, A., Granhag, P. A., & Porter, S. B. (2010). Pitfalls and opportunities in nonverbal and verbal lie detection. *Psychological Science in the Public Interest*, **11**, 89–121. doi:10.1177/152910061039086.
- Zuckerman, M., DePaulo, B. M., & Rosenthal, R. (1981). Verbal and nonverbal communication of deception. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 14, pp. 1–59). New York, NY: Academic Press.