

## ORIGINAL ARTICLE

**Content in Context Improves Deception Detection Accuracy**J. Pete Blair<sup>1</sup>, Timothy R. Levine<sup>2</sup>, & Allison S. Shaw<sup>2</sup><sup>1</sup> Department of Criminal Justice, Texas State University, San Marcos, TX 78666, USA<sup>2</sup> Department of Communication, Michigan State University, East Lansing, MI 48824, USA

*Past research has shown that people are only slightly better than chance at distinguishing truths from lies. Higher accuracy rates, however, are possible when contextual knowledge is used to judge the veracity of situated message content. The utility of content in context was shown in a series of experiments with students (N = 26, 45, 51, 25, 127) and experts (N = 66). Across studies, average accuracy was 75% in the content in context groups compared with 57% in the controls. These results demonstrate the importance of situating judges within a meaningful context and have important implications for deception theory.*

doi:10.1111/j.1468-2958.2010.01382.x

The power of the situation in determining human behavior is among the most important lessons of social science (Ross & Nisbett, 1991). A second and related lesson is the human tendency to underestimate the importance of situations (Ross & Nisbett, 1991). It follows from this second principle that the first likely needs to be repeatedly relearned and reapplied.

The literature on deception detection is an excellent example of an area in social science where the power of the situation is underappreciated. Past theory and research on deception has been characterized by a myopic focus on the internal psychological states and corresponding nonverbal behaviors of liars and has failed to adequately consider the situation and context in which truths and lies are told. As a result, deception research has been looking for cues to deception in the wrong places and the existing literature presents a distorted view of people's ability to correctly assess if others are lying.

Consider the aggregate findings from the past 40 years of deception detection experiments. Meta-analysis of more than 200 experiments finds that people are significantly better than chance at distinguishing truths from lies, but seldom much better than chance (Bond & DePaulo, 2006). Across studies, average accuracy is 54%—just 4% better than the .50 base chance level. Further, findings over the years have been amazingly consistent from study to study. Accuracy results from individual

---

Corresponding author: Timothy R. Levine; e-mail: levinet@msu.edu

studies are normally distributed around this across-study average, and the majority of studies report results within just 6 percentage points of the across-study average (Bond & DePaulo, 2006).

Contrast this conventional understanding of the deception detection literature with the findings from one of the most famous deception detection experiments in the history of social science. Asch (1956) had subjects judge the length of a series of lines after exposure to verbal reports of the judgments of others. On critical trials, research confederates always gave incorrect answers. Yet, despite the obvious falsity of the confederates' answers, virtually none of Asch's subjects concluded that they were being lied to. As Asch observed: "Instances of suspicion were rather infrequent. One would expect distrust to grow as the majority continued to err. [But] Before his [the subject] suspicions had the opportunity to take root, he had unwittingly come to doubt himself and to seek explanations in other directions" (p. 29). "Most subjects did not suspect that the majority of judgments were not genuine. Suspicion at times occurred only as an [sic] hypothesis which, like many others, was rejected" (p. 31).

Similarly, apparently none of Milgram's (1969) obedience subjects pegged the victim as an actor who was blatantly lying about being shocked. In our own experiments, some of which are reported here, after more than 300 individual sessions, the confederate was correctly identified as an imposter only once. So, why do subjects in deception detection experiments consistently do slightly better than chance while subjects in experiments involving deception, but not about deception, achieve accuracy for deception that approaches zero?

A key aspect of the answer, we believe, lies in the situation, and more specifically how people construe the situation. Deception experiments are, after all, about deception. Subjects who are exposed to odd, unexplainable behavior and who know they are in a deception detection experiment infer deceit (Bond et al., 1992). This is a perfectly reasonable inference given the situation. Alternatively, subjects exposed to similar sorts of odd behaviors but who do not know that they are in a deception study almost never infer deception from odd behaviors, inferring that the people are just odd rather than deceptive (Levine et al., 2000). Even in the absence of odd behavior, subjects in deception detection experiments still know (or can infer) that they need to guess lie some of the time and consequently their guesses are correct at rates that (mostly) approximate chance. Indeed, chance models can predict accuracy in these experiments quite well when it is known what proportion of the time lie guesses are made (Levine, Kim, Park, & Hughes, 2006). By contrast, there was little reason for the subjects in Asch (1956) or Milgram (1969) to suspect deceit from the confederate. People lie when situational forces make the truth problematic (Levine, Kim, & Hamel, 2010) and people infer that this is just what other people do (the projected motive model; Levine, Kim, & Blair, 2010).

We believe that the situation not only tells people when they need to suspect deception but also can provide useful clues as to which messages might be deceptive and honest within situations. One of the simplest ways to detect deception is to compare the content of what someone says with what is already known. Statements

that explicitly contradict other information that the receiver possesses will be taken to indicate deception when the receiver is primed to detect deception. For example, if a sender claims to have been at home on a given night, but the receiver was told by a trusted source that she saw the sender out a restaurant on the night in question, it is likely that the sender's statements will be flagged as deceptive.

The utility of context in guiding deception judgments is not limited to the explicit contradictions of existing contextual information. Context can also provide information about what is normal or possible in a given situation. This normative information can come in a wide variety of forms. These include, but are not limited to: knowledge about the sender's normal activities; beliefs about how a given situation typically unfolds; the laws of physics and nature; and, information about how people normally perform in a given situation. When normative information of these sorts is present, the sender's statements can be considered in light of this knowledge. If the sender's statements or performance are implausible given this normative information, the statements can then be flagged as potentially deceptive. If a person who is less than 5-feet tall claims that she can slam dunk a basketball on a regulation basket ball hoop, the receiver's normative knowledge makes it quite likely that this statement will be judged untruthful.

Context can also provide idiosyncratic information that does not directly contradict what the suspect says, or provide normative information, yet aids in the detection of deception. This type of information can be considered to be analogous to circumstantial evidence in a criminal case. Idiosyncratic knowledge does not directly reveal deception; rather, it points to a higher probability of deception occurring. Consider a situation in which a number of shortages have occurred at a bank. The shortages stop when one of the employees goes on vacation and begin again when the employee returns. When questioned about the shortages, the employee's statements may not explicitly contradict what the interviewer knows, normative information may not suggest that the employee's statements are implausible, but idiosyncratic knowledge about the starting and stopping of the shortages may still cause the interviewer to believe that the employee's statements are deceptive. The sender's statements are still important in this context as they may serve to explain away the idiosyncratic information. The sender could, for example, point to another employee who took a vacation at the same time as the sender or point out that the bank was closed during her vacation, thus making shortages impossible during that time.

These three types of knowledge about the situation and sender form a group of cues that we refer to as "content in context." It is our belief that lies that are told within a meaningful context can often be detected on the basis of these "content in context" cues. For example, all the lies by the Asch (1956) confederates could have been easily distinguished from honest responses had Asch's subjects known the nature of the experiment. On the other hand, the decontextualized nature of the lies in most deception detection experiments makes message content of little use and consequently accuracy is only slightly better than chance. There are, of

course, situations outside the lab where the context is of little use, but social actors typically have some quasi-accurate knowledge of context. Without this knowledge, communication would be much less efficient.

We contend that the existing deception detection literature tells us only about people's ability to detect deception when (a) people know deception is to be expected and (b) when the context is of little help in distinguishing truths from lies. In such cases, people do slightly above chance. Noncontent cues to deception can have some very limited diagnostic utility, but are sufficiently unreliable (DePaulo et al., 2003; Sporer & Schwandt, 2006, 2007) to allow accuracy at rates much above chance (Levine, 2010). Consequently, in such situations, individual differences (Aamodt & Custer, 2006; Bond & DePaulo, 2008) and training (Frank & Feeley, 2003) make little difference, and accuracy is never very much above chance. But, this tells us little about deception detection outside the lab because this is not how people usually detect lies (Park, Levine, McCornack, Morrison, & Ferrara, 2002).

In contrast, the use of deception in research about nondeception topics tells us that in the absence of contextual information prompting suspicion, people tend to believe others even when the content in context suggests deception (e.g., Asch, 1956). In the case of Milgram (1969), how likely is it that a Yale psychologist would torture subjects to near death with electric shocks? It is unthinkable. Nevertheless, people found the situation believable and the consistent believability of research confederates is probably a more ecologically valid conclusion than those of the typical deception detection experiment.

What research has failed to study is situations where people are trying to detect deception and where they have some limited, but useful knowledge of the situation. Outside the deception lab, lies are typically told in some context and situational background knowledge is often both known and useful. Police have crime scenes and reports from witnesses. Relational partners have relational histories and knowledge of their partner's routines. Political constituents know past records or can fact check on the Internet. Juries can compare testimony from witnesses to the other facts of the case. Employers can give drug tests and perform criminal records checks. In such situations, we believe that the content in context can have a degree of diagnostic utility that surpasses the utility of nonverbal behaviors stemming from internal states believed to be associated with deception.

In the series of experiments reported here, we test the hypothesis that humans can accurately detect deception at rates well above chance. We believe that such an outcome is possible through reliance on content in context rather than observation of nonverbal behavior. Support for this hypothesis would stand in stark contrast to both the 200+ experiments that preceded us and the leakage-based theoretical accounts that dominate past and current understandings of deception detection (e.g., Ekman & Friesen, 1969). Support for our hypothesis would once again demonstrate both the impressive power of the situation and the impoverished and distorted view that results from failing to take the nuances of context into account.

## Study 1

The truths and lies used in Studies 1, 2, 3, 6, and 7 were selected from a database of honest and deceptive videotaped interviews created by Levine (2007) that was funded by the National Science Foundation. Briefly, the participants interviewed in these videotapes were told that the focus of the research was on teamwork, that the study involved playing a trivia game with a partner for a cash prize, and that after the game they would be interviewed about their experience in the study. During the trivia game, all participants were given an opportunity to cheat when the researcher was called out of the room, and their partner, a research confederate, suggested cheating. Participants decided for themselves whether to cheat, and if they cheated, whether to lie about it during the interview phase. In the videotapes used in the current experiments, all interviewees denied cheating. The denials of the noncheaters were honest and the denials of the cheaters were lies. Because it is known who actually cheated, ground truth is certain in the tapes. Further, all lies in the cheating tapes were unsanctioned (i.e., liars were not instructed to lie, nor was there any indication that lying was acceptable in the situation until debriefing), and the situation was one of potential high stakes because the lies were about cheating in federally funded research at a university. In addition, because a cash prize was offered, cheating was analogous to stealing. Subjects in the current experiments watched a series of these videotaped interviews and were asked to judge if each interviewee was an honest noncheater or a lying cheater.

In the cheating tapes, some potentially useful, contextual information regarding who cheated and who did not was the knowledge that the trivia questions were very difficult. Few teams had more than two answers correct without cheating. For noncheaters, the confederate contributed one correct answer, and few participants knew more than one answer on their own. The interview questions asked about the participant's performance, and participants knew their own scores. Thus, listening to content about performance has potential efficacy in this context. People who did poorly in the game were usually noncheaters while teams that have several answers correct likely cheated. The "content in context" cue in these studies, then, dealt with normative information.

## Method and results

As an initial test of our hypothesis, 26 undergraduate students (24 females, 2 males; mean age = 21.2 years,  $SD = 0.78$ ) enrolled in an upper division communication class at a large state university in the midwestern United States watched a series of 12 videotaped interviews (six honest and six lies) randomly selected from the tapes created by Levine (2007) as described previously. Participants were asked to judge the person interviewed in each tape as either an honest noncheater or as a lying cheater, and these judgments were scored for accuracy. After watching and judging each of the 12 tapes once, the participants were informed that the questions in the trivia game were difficult and that they should pay special attention to what the interviewees said about their performance. They then watched and rated the tapes a second time.

The results were not consistent with our hypothesis. Accuracy in the two groups was not statistically different, paired  $t(25) = -1.00$ ,  $p = ns$ ,  $d = .17$  and the means are provided in Table 1.

## Study 2

The results of Experiment 1 failed to support our prediction that the content in context is useful in enhancing deception detection accuracy. Alternative explanations for the results are that our thinking is simply wrong, that the content in context can be useful, but the tapes simply did not contain sufficient content, that the context information induction was too weak, or that the subjects simply did not take the experiment seriously and consequently failed to use available content because of a lack of effort. Study 2 involved a content analysis of the Levine (2007) tapes to ascertain if diagnostically useful content was present.

### Method and results

To examine the presence of the contextually useful content, two coders viewed each of the 100 videos involving denials in the Levine (2007) database. The coders were blind to the veracity of the interviewees and worked independently. Two coding schemes were used; one with five categories and the other with six options. In both of these schemes, the “high” categories referred to explicit statements by the interviewee’s suggesting that they scored more than two answers correct on the trivia test, and the low categories were used for statements suggesting that the interviewee’s score was two or less on the trivia test. It should be noted again that the participants were never directly asked to state their scores during the course of the interviews. In the five-option coding, an interview was rated as explicitly high if the participant indicated that his or her score on the trivia test was three or more. Tapes were coded as implicitly high if the participant implied his or her score was high, but did not give enough information to determine if the value of the score was three or more. The coders indicated no judgment if the participant did not give enough information to make a determination. The implicitly low category was used if the participant implied that his or her score was low, but did not give enough information to determine that the score was two or less. Videos were scored as explicitly low if the participant explicitly mentioned that his or her score was two or less.

When the coders were unable to make a judgment, they were asked to indicate a trend as to whether the participants’ answers were more consistent with a high or low score on the trivia test. This, in essence, forced a judgment on each video and produced a 6-point scale with three values on the high trivia score side and three on the low trivia score side.

Krippendorff’s alpha ( $K\alpha$ ) was calculated as an indication of intercoder reliability (Krippendorff, 2004). Reliabilities were .89 and .86 for the five- and six-option coding, respectively, indicating that the tapes were coded with acceptable reliability.

Table 1 Study Results

Study-Condition	N	Total Accuracy (SD)	Truthful Accuracy (SD)	Deceptive Accuracy-Hits (SD)	False Alarms (SD)	Truth Bias (SD)
Study 1						
Control	26	59% (14%)	67% (19%)	50% (19%)	33% (19%)	59% (14%)
Context	26	56% (20%)	64% (20%)	48% (18%)	36% (20%)	53% (17%)
Study 3						
Control	45	61% (15%)	74% (19%)	57% (17%)	36% (19%)	54% (11%)
Context	45	77% (14%)	79% (16%)	85% (17%)	31% (16%)	42% (09%)
Study 4						
Control	26	33% (16%)	36% (23%)	40% (25%)	73% (23%)	43% (18%)
Context	25	80% (13%)	76% (17%)	95% (17%)	34% (17%)	35% (10%)
Study 5						
Control	12	42% (22%)	50% (30%)	33% (25%)	50% (30%)	58% (16%)
Context	13	69% (23%)	69% (33%)	69% (25%)	31% (33%)	50% (18%)
Study 6						
No context—no performance	34	62% (11%)	73% (20%)	51% (22%)	27% (20%)	61% (19%)
Context—no performance	31	65% (13%)	70% (17%)	59% (18%)	30% (17%)	56% (17%)
No context—performance	30	59% (19%)	60% (26%)	58% (23%)	40% (26%)	51% (16%)
Context—performance	32	73% (14%)	67% (17%)	80% (19%)	33% (17%)	43% (11%)
Study 7						
Control	32	69% (12%)	79% (17%)	58% (22%)	21% (17%)	61% (15%)
Context	34	72% (16%)	73% (22%)	70% (23%)	27% (22%)	51% (16%)

Table 1 Continued

Study-Condition	N	Total Accuracy (SD)	Truthful Accuracy (SD)	Deceptive Accuracy-Hits (SD)	False Alarms (SD)	Truth Bias (SD)
Study 8						
Control	32	34% (13%)	29% (18%)	39% (15%)	71% (18%)	45% (10%)
Context	34	81% (15%)	74% (16%)	89% (20%)	26% (16%)	42% (10%)
Study 9						
Control	32	68% (20%)	77% (28%)	59% (20%)	23% (28%)	58% (14%)
Context	34	75% (25%)	82% (30%)	68% (30%)	18% (30%)	57% (17%)
Weighted averages						
Control	176 <sup>a</sup>	75% (SE = 1%)	74% (SE = 1%)	78% (SE = 1%)	26% (SE = 1%)	47% (SE = 1%)
Context	237 <sup>a</sup>	57% (SE = 1%)	63% (SE = 1%)	52% (SE = 1%)	37% (SE = 1%)	54% (SE = 1%)

Note: Ns denoted with the superscript “a” denote the number of unique judges. The same experts made judgments in Studies 7, 8, and 9.



**Table 2** Coding Judgments by Status for the 5-Point Scale

Judgment	Truthful		Deceptive		Total	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
Explicit high	4	5.1	10	45.5	14	14.0
Implied high	2	2.6	3	13.6	5	5.0
No judgment	29	37.2	7	31.8	36	36.0
Implied low	7	9.0	0	0.0	7	7.0
Explicit low	36	46.2	2	9.1	38	38.0
Total	78	100.00	22	100.00	100	100.00

The coding results are presented in Tables 2 and 3. Both sets of results suggested that coded content was useful in distinguishing honest from dishonest denials. With the 5-point coding, 13 (59.1%) of the deceptive participants gave answers that were coded as indicating a high score, and 42 (55.2%) of the truthful participants gave answers that were coded as indicating a low score. For the deceptive participants, the coders were unable to assign a score in 7 (31.8%) cases, and the coders were unable to assign a score for the truthful participants in 29 (37.2%) cases. It was unusual for a deceptive participant to be coded as having a low score, and this occurred in only two (9.1%) cases. The reverse was true for the truthful participants; only six (7.7%) were coded as having a high score. Given the low expected frequencies in some cells, a chi-squared test was not conducted. Instead, the explicit and implicit categories for both high and low scores were collapsed to produce a 3 (high–no judgment–low)  $\times$  2 (truthful–deceptive) table. The test of this table was significant ( $\chi^2_{(2)} = 32.08, p < .001$ ) and revealed a large effect size for status on cue presence (Cramers  $V = .57$ ). Similar conclusions were obtained in the six-option coding. When collapsed into a 2 (high–low score)  $\times$  2 (truthful–deceptive status) table, 64 (82.1%) of the truthful and 18 (81.8%) of the deceptive participants were coded as exhibiting a content in context cue that was consistent with their status. The test of this table was significant ( $\chi^2_{(1)} = 32.17, p < .001$ ) and revealed a large effect size for status on cue presence ( $\phi = .57$ ).

The results of the coding in Study 2 revealed that the content in context cue could be coded reliably, and that the participants displayed the cue in a manner that was consistent with their truthful or deceptive status. As was illustrated by the 82% correct classification rate, this was particularly true when the forced-choice coding was considered.

It appears that the cue can be subtle. When not forced to make a call one way or the other, the coders indicated that they could not make a decision in about one-third of the cases. The analysis of the forced-choice decisions, however, indicated that the coders selected the correct direction in 72.4% of the truthful cases and 71.4% of the deceptive cases that they initially identified as inconclusive. The apparent subtlety of the content cues may, in part, explain the failure of Study 1. If the task requires a

**Table 3** Coding Judgments by Status for the 6-Point Scale

Judgment	Truthful		Deceptive		Total	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
Explicit high	4	5.1	10	45.5	14	14.0
Implied high	2	2.6	3	13.6	5	5.0
Trend high	8	10.3	5	22.7	13	13.0
Trend low	21	26.9	2	9.1	23	23.0
Implied low	7	9.0	0	0.0	7	7.0
Explicit low	36	46.2	2	9.1	38	38.0
Total	78	100.00	22	100.00	100	100.00

motivated and attentive judge, and if the participants in Study 1 were not paying sufficiently close attention to either the context information induction, the content in the interviews, or both, then the failure of Study 1 can be dismissed as attributable to a lazy subject artifact. Study 3 tested if motivated subjects might provide more supportive results.

### Study 3

#### Method and results

Study 3 involved the same research design as the first experiment except that an incentive was added to encourage participants to take the research task seriously and slightly more detailed contextual information was provided. Forty-five (46% females, mean age = 23.89,  $SD = 3.37$ ) students in criminal justice classes at a large state university in the Southwestern United States volunteered to participate in return for course credit. To motivate the students to perform well, participants were told that they would receive maximum course credit if they were able to correctly identify more than half of the participants as either truthful or deceptive. If they failed to reach this level, they would receive only half credit. Each participant watched and judged the same 12 interviews used in Study 1. After the first set of judgments was completed, the participants were given information on the context in which the videotapes were made. They were told that the trivia test was extremely difficult. The questions were so difficult, in fact, that the people in the videos were rarely able to have more than two correct if they did not cheat. Following this contextual information, the participants were instructed to watch the videos again and indicate whether they thought that the people in the videos were either truthful or deceptive.

The mean number of correct judgments on the first viewing of the videos was 61% ( $SD = 15\%$ ). The mean number of correct judgments on the videos after the contextual information was provided was 77% ( $SD = 14\%$ ). This difference was significant, paired  $t(44) = 6.06$ ,  $p < .001$ ,  $d = 1.10$ . The mean number of correct judgments on the first and second runs were both significantly above chance:

one-sample  $t(44) = 4.90, p < .001, d = .76$  and one-sample  $t(44) = 13.04, p < .001, d = 1.86$ , respectively.

Participants hit (correctly identified deceptive materials as deceptive) on 57% ( $SD = 17\%$ ) of the deceptive materials on the pretest and 85% ( $SD = 17\%$ ) of the deceptive materials on the posttest. This difference was significant and indicative of a large effect size, paired  $t(44) = 7.70, p < .001, d = 1.59$ . Participants had false alarms (identified truthful materials as deceptive) on 36% ( $SD = 15\%$ ) of the truthful materials on the pretest and 31% ( $SD = 14\%$ ) on the posttest. This difference was not significant, paired  $t(44) = 1.42, p = .16, d = .23$ .

The results of Study 3 suggested that when participants were made aware of useful contextual information, and when participants were motivated to attend to that information, improved deception detection accuracy was achieved. Raw accuracy in the content-informed condition was 77% compared with 56% and 59% accuracy in Study 1 and 54% in the Bond and DePaulo (2006) meta-analysis. Further, most of the gain from utilizing content in context came from an increased ability to identify deceptive messages.

The repeated measures design of the study, however, made it possible that the improvement from the first set of judgments to the second was simply the result of having seen the videos on two occasions. Study 4 was conducted to rule out a “mere practice” alternative explanation as well as to test if content in context reasoning might generalize to lies in different situations.

## Study 4

### Method and results

Fifty-one students (64% males, mean age = 23.1,  $SD = 3.79$ ) in upper division criminal justice classes at a large state university in the southwestern United States volunteered to participate in the study in exchange for extra credit. Each participant viewed a sample of 10 true or false videotaped confessions to mock crimes. Five of the confessions were false, and each of the 10 confessions was to a different mock crime with different case facts.

Participants were randomly assigned to either a context-informed or a control condition. In both conditions, participants were told that they were investigators handling robbery cases and that a suspect had confessed to a robbery that they were investigating. Their task was to decide whether or not the confession was true. Participants were also informed that any number of the confessions could be true or false. The participants in the control condition simply viewed the video of each confession and then indicated whether they thought that the confession was true or false. In the contextual information condition, the participants were given a “case file” that contained a mock-up of a typical preliminary investigation report. This information included a description of the victim, where the crime occurred, what weapon (if any) was used, and what was stolen. This information provided a context in which the sender’s statements could be judged against the statements of the victim

and police officer. As such, this scenario dealt with the explicit contradiction of the knowledge possessed by the receiver class of content in context cues. After the judges reviewed this information, they watched the videos of the confessions. Following each video, the participants were asked to indicate whether the confession was true or false.

As Table 1 indicates, those in the contextual information group were correct on 80% of their judgments ( $SD = 13\%$ ). Participants in the control group were correct on only 33% of their judgments ( $SD = 16\%$ ). This difference was significant and the size of this effect was large,  $t(47.78) = 11.50$ , equal variances not assumed,  $p < .001$ ,  $d = 3.2$ .

The contextual information group hit on 95% ( $SD = 17\%$ ) of the deceptive materials and had false alarms on 34% ( $SD = 17\%$ ) of the truthful materials. The control group hit on 40% ( $SD = 25\%$ ) of the deceptive materials and had false alarms on 73% ( $SD = 23\%$ ) of the truthful materials. Both of these differences were significant and indicated large effect sizes,  $t(43.37) = 9.24$ , equal variances not assumed,  $p < .001$ ,  $d = 2.7$  for hits and  $t(46.20) = 6.95$ , equal variances not assumed,  $p < .001$ ,  $d = 2.0$  for false alarms.

The results of Study 4 replicated those of Study 3 with an independent groups design and a different set of situations and messages. Access to contextual information improved overall accuracy and hit rates while reducing false alarms. These results rule out the mere practice explanation for previous findings. The results further suggested that there may be situations where below chance accuracy is possible when crucial contextual information is withheld.

## Study 5

Study 5 sought to replicate the results of Studies 3 and 4 using materials from a nonresearch setting. The lies in the previous studies involved either students cheating on an experimental task or confessing to a mock crime. Research-created lies are of ethical necessity lower in stakes than some nonresearch lies. Study 5 applied the content in context logic to the detection of some high stakes lies.

### Procedure and materials

Four cases were selected from the corpus of videos first reported by Horvath, Jayne, and Buckley (1994). Each of these cases involved an employee theft of money from an employer. The cases were selected because they each contained two and only two primary suspects. One of the suspects was deceptive (guilty) and one was truthful (innocent) in each case as established by the evidence-corroborated confession of the deceptive party (see Horvath et al., 1994, for more information on the videos). From the four cases, two truthful and two deceptive suspects were randomly selected and placed in random order on a video; thus, participants saw only one of the two prime suspects from each case.

Twenty-five students (60% females, mean age = 27.6,  $SD = 9.7$ ) at a large university in the southwestern United States were randomly placed into either the context ( $n = 13$ ) or the control group ( $n = 12$ ). Participants in the context group received a case file that contained a brief summary of the background of the prime suspects and the general scenario regarding the missing money. They were instructed to read the case file prior to viewing the relevant video. Participants in the no-context group did not receive this information. Both groups viewed the videos and rendered dichotomous truth–deception judgments.

### Results and discussion

Participants in the context group were more accurate overall (69%) and scored more hits (69%) on the deceptive materials than those in the control group [control group accuracy = 42 and 33%,  $t(23) = 3.03$ ,  $p < .01$ ,  $d = 1.2$  and  $t(23) = 3.59$ ,  $p < .01$ ,  $d = 1.4$ , respectively]. Groups did not differ in terms of false alarms, truth accuracy, or truth bias. These findings replicated the findings of Studies 3 and 4 using an independent groups design and set of “real life” materials. Once again, context improved accuracy, particularly deception detection accuracy.

The suspects in these videos did not explicitly contradict the information in the case files at any time during their interviews. Although there is the possibility that the case file provided some normative information regarding the propensity of suspects to commit the theft (such as age, job title, and length of employment), the files were awash in idiosyncratic information about the activities of the suspects around the time of the theft, and we believe that it is this information that provided the relevant content in context cues. We cannot, however, eliminate the possibility that it was normative contextual information that was responsible for the results.

### Study 6

The results so far are consistent with the utility of the content in context perspective. Study 2 showed the existence of useful content in a sample of real truths and lies, and Studies 3, 4, and 5 demonstrated that substantially better than chance accuracy was possible when motivated judges had relevant knowledge of context. Study 6 assesses the independent necessity of the two key elements of the content in context logic—context-relevant content must be available in the communication under scrutiny and the judge must know the relevant contextual information in order to draw a correct inference from the content. Study 6 tested the hypothesis that both these elements are necessary and jointly sufficient to produce improved accuracy in motivated judges.

### Design and materials

The participants were 127 undergraduate students (53% females, mean age = 19.8,  $SD = 2.1$ ) who were part of a departmental research pool at a large state university in the midwestern United States. Participation fulfilled a class research requirement.

As an extra incentive, participants who achieved at least 75% correct were paid \$5 and \$20 was paid anyone who achieved 100% accuracy.

A  $2 \times 2$  independent groups experimental design involved crossing the provision of context information (context informed and context uninformed) with access to contextually relevant content (performance questions asked, not asked). The context information induction was similar to previous experiments. In the informed condition, participants were told that the questions were difficult, and that scores above two often indicated cheating. In the information withheld condition, this context information was not provided.

To manipulate the access to content, the Levine (2007) cheating videotapes were edited. Each full interview consisted of three parts: some background questions asking about prior teamwork experience and about experience playing trivia games, questions about performance in the current game, and finally, questions asking directly about cheating and honesty. The useful content in context information was typically provided in the second part in response to the performance questions while the denial of cheating occurred in the final part of the interview under direct questioning. The tapes were edited such that all contained the background questions but either the performance questioning or the direct questioning about cheating was omitted. Participants in Study 6 were randomly assigned to view either tapes containing direct questioning or performance questioning and were either provided context information or not. All viewed and rated tapes of the same 12 interviews, half of which were cheaters. The judgments were scored for accuracy.

### Results and discussion

Accuracy in the critical context informed cue present cell was 73%. Accuracies in the other three cells were 61%, 58%, and 65%. The test of the *a priori* contrast of +3, -1, -1, -1 was statistically significant,  $t(123) = 4.01$ ,  $p < .001$ , and Tukey B tests showed that the critical cell differed from the other cells at  $p < .05$ , and the other three cells were not significantly different from each other. Hits, false alarms, and truth bias in each cell are provided in Table 1.

Consistent with the content in context explanation, the results of Study 6 showed that both knowledge of context and context-relevant content are needed for enhanced accuracy. In the absence of either of the elements, lower levels of accuracy are observed. Consistent with previous findings, the results of Study 6 also showed the most notable gains in the correct identification of lies.

### Studies 7, 8, and 9

The participants in the previous experiments were college students. Studies 7, 8, and 9 sought to examine whether or not the previously reported results would extend to nonstudents. The participants in the next three experiments were experienced polygraph examiners and investigators who worked for a federal agency. They could, therefore, be considered experts at deception detection.

### Participants and procedure

Sixty-six participants (59% females, mean age = 31.9,  $SD = 9.9$ ) completed the studies. The participants were randomly divided into context and control groups and the groups were placed in separate rooms. Thirty-four were in the context condition and 32 were in the control condition. Both groups watched the same cheating (Study 7), confession (Study 8), and criminal videos (Study 9) that had been shown to the participants in Studies 1, 3, 4, and 5. Participants in both conditions were given inductions that were similar to those in Studies 3, 4, and 5, the only difference being that the course credit and cash motivational inductions were not included.

### Study 7

#### Results and discussion

Overall accuracy was not different between the two groups, although the means were in the predicted direction [context informed  $M = 72\%$ ,  $SD = 16\%$ , context uninformed  $M = 69\%$ ,  $SD = 12\%$ ,  $t(65) = 0.90$ ,  $p = .37$ ,  $d = .21$ ]. Significant differences were observed, however, in hits [context informed  $M = 71\%$ ,  $SD = 23\%$ , context uninformed  $M = 58\%$ ,  $SD = 22\%$ ,  $t(65) = 2.48$ ,  $p < .02$ ,  $d = .58$ ; see Table 1].

Although support was not obtained from the significant test for total accuracy, the lack of difference appears more a function of unusually high accuracy in the control group (68.5%) than poor performance in the context-informed group (71.7%). It is possible that the experienced investigators could have inferred the contextual relevance without explicit information. Alternatively, they may have picked up on other types of useful information. The relative advantage of content in context for lies, however, was again observed.

### Study 8

#### Results and discussion

Participants in the context group ( $M = 81\%$ ,  $SD = 15\%$ ) were more accurate than those in the no-context group ( $M = 34\%$ ,  $SD = 13\%$ ) and this difference suggested a large effect size for the presence of context,  $t(64) = 13.67$ ,  $p < .001$ ,  $d = 3.4$ . Participants in the context group were also more accurate in truth and lie accuracy, and exhibited fewer false alarms than those in the control group. Consistent with our general hypothesis and the rest of the studies, the presence of contextual information improved the ability of the participants to detect deception.

### Study 9

#### Results and discussion

Overall, participants in the context group were not more accurate than those in the control group,  $t(64) = 1.26$ ,  $p = .21$ ,  $d = .31$ . The groups were also not significantly



different in terms of hits, false alarms, truth accuracy, or truth bias. The means of each of the accuracy measures were in the predicted direction and it appears that it was the high accuracy of the control group (68%) that produced the nonsignificant results. It could be that the experienced investigators inferred contextual relevance without being given explicit information, or that they picked up on other types of useful information. The study indicated, nonetheless, that above chance accuracies are possible when participants have access to contextual information.

## Study 10

Study 10 presents a mini meta-analysis of all of the overall accuracy results of the previous studies (with the exception of the content analysis results presented in Study 2) in this article. The eight context-informed conditions reported in this article consisted of a total of 176 judges who made a total of 2,422 judgments. Accuracy ranged from a low of 56% to a high of 81%. Standard deviations ranged from 13% to 25%. The weighted mean average of these accuracy estimates was 75% with a standard error of approximately 1%. This resulted in a 95% confidence interval ranging from a low of 73% to a high of 77%, and suggested that the participants in the context conditions of this article were about 20% more accurate than the participants in the DePaulo et al. (2006) meta-analysis. This accuracy places the accuracy of the participants reported in this article clearly to the extreme right of the accuracy distribution reported by Bond and DePaulo (2006). Weighted mean averages and standard errors are reported for the other context present accuracy measures at the bottom of Table 1.

The 10 control condition accuracies reported in this article consisted of a total of 237 judges who made 3,132 judgments. Accuracy ranged from a low of 33% to a high of 69%. Standard deviations ranged from 12% to 22%. The weighted mean average of these accuracy estimates was 57% with a standard error of approximately 1%. This produced a 95% confidence interval that ranged from a low of 55% to a high of 59%. This finding suggested that the accuracy of the participants in the control conditions was slightly higher than the 54% reported by Bond and DePaulo (2006). Weighted mean averages and standard errors are reported for the other control group accuracy measures at the bottom of Table 1.

## Overall discussion

It has become nearly axiomatic in the deception detection literature that people are able to detect deception at a level that only slightly exceeds chance (Levine, Park, & McCornack, 1999; Vrij, 2000). The experiments reported here demonstrated that accuracies significantly higher than 54% can be consistently achieved when motivated judges are given access to meaningful contextual information. The mean accuracy of the contextual groups in our studies was a full 21% higher than found by the Bond and DePaulo (2006) meta-analysis. Providing accurate contextual information improves deception detection accuracy.



This finding may seem obvious, but in 40+ years of active deception research, the role of context has been largely ignored. Research has considered the impact of some contextual features such as the type of motivation induction or whether the lie is about a transgression or not, but this is performed under the premise that these contextual features will impact internal states and leakage-type displays. As we noted in the beginning of this article, this approach has produced very modest results in the last 40 years and is inconsistent with how people detect deception outside of deception detection experiments (Park et al., 2002).

The content in context approach taken here has produced exceptionally large effect sizes across samples and scenarios. Even the small amount of contextual information presented here produced accuracy swings of up to 47% and explained as much as 72% of the variance in accuracy (Study 4).

Not all situations will produce meaningful content in context. We conceptualize deception contexts as being arrayed along a continuum. At the origin of the continuum are those contexts that do not provide the deception judge with information that can assist his or her judgments. At the extreme right-hand side of the continuum are those contexts that provide the judge with so much information as to make veracity judgments trivial. A situation wherein a judge is asked to determine whether or not a nurse is lying about the content of a video would lie near the origin. Situations wherein a 6'10", heavily muscled individual with a deep voice, thick beard, and other masculine features claims to be female would be located toward the right side of the continuum.

Although no research to date has explicitly examined the actual distribution of deception detection contexts along this continuum, we believe that outside the controlled conditions of the deception lab, a substantial number of contexts are located to the right of the origin of the continuum. That is, many (if not most) everyday deception contexts provide the judge with at least some meaningful contextual information. As noted previously, it is often this contextual information that prompts the judge to even consider the possibility of deception. By focusing exclusively on scenarios that are devoid of this information, the traditional detection of deception experiments tell us little, if anything, about the ability of judges to detect deception in many common contexts.

Our work is closely tied to the approaches of Brunswick (1955) and Gigerenzer (2008), who argue that to study the functioning of a psychological process, both the process and the environment in which the process operates must be considered. This approach is grounded in a Darwinian understanding of how psychological processes develop. The belief is that processes are adapted to the environments in which they were developed. Processes that are functional and adaptive are retained and those that are not are selected out. The metaphor of a pair of scissors is often utilized to explain this approach (Simon, 1990). The first blade on the pair of scissors represents the cognitive, affective, or behavioral process and the second represents the environment. The functioning of the pair of scissors can only be understood when both blades are considered. Considering only one blade or the other will not reveal how paper is cut.

The study of deception detection up to this point has been focused almost exclusively on the first blade of the scissors. There have been deception studies in which the effects of varying environments have been considered, but these studies have been guided by theories positing that the effects of environment are only important in that they cause more or less leakage to occur. Consider the deception cues meta-analysis conducted by DePaulo et al. (2003). The moderators that were considered (transgressions and motivation) were explored because the existing theories posited that these changes in environmental conditions would moderate deceptive displays through some variety of leakage.

Consistent with these dominant leakage theories, researchers have designed their experiments to have tight internal validity so that the causal claims of their theories can be assessed. The focus of the dominant theories requires that anything extraneous to leakage be eliminated. This makes sense for theories in which the primary mechanisms are deception-linked thoughts and feelings that produce leakage and nonverbal deception cues. These concerns, however, have the unfortunate consequence of shifting attention away from other mechanisms that may be useful in understanding deception detection and in creating contexts that differ in important ways for deception detection outside the lab.

In this article, we have presented a set of conditions in which the second blade of the scissors (the environment) is allowed to function in concert with sender behavior. Nonverbal leakage in the studies presented here is constant across conditions because only contextual information was varied (except in Study 6). The results of the tests presented here are overwhelming. When judges were asked to make deception judgments with some meaningful contextual information, they performed significantly better than chance and significantly better than 40+ years of research suggests they would. Clearly, knowledge of the environment in which deception occurs facilitates accurate deception judgments beyond what is possible based on observations of nonverbal leakage. Given the large amount of variation explained by the differences in environments (context), deception theories will be enhanced by explicitly recognizing the impact of context.

Gigerenzer has developed a model for explicitly examining psychological processes and their functioning in the environment through his work on heuristics. First, the functioning of the process is identified. Second, the functioning of the process in a variety of ecologically sound environments is evaluated to assess the limits of the process. Third, this information can be utilized to alter the environments in which people operate in a way that allows the process to function successfully (Gigerenzer, 2008). Gigerenzer and his colleagues have had substantial success in identifying environments in which heuristics that were considered to be inaccurate actually function well (see Gigerenzer, 2008, pp. 14–15 for a list). We believe that this model provides a method for moving our understanding of deception detection forward.

It appears that in contexts where the detection of deception is critical, practitioners have already taken steps to maximize the availability of contextual information of the types that we considered in this article. Criminal investigators interview suspects

after they have collected information from trusted sources; intelligence agencies cross-check information from human sources against other collected information; banks ask for verification of employment and income before issuing loans; and retailers require a valid card, which is checked against an electronic database, before selling items on credit. In these ways, it appears that the practitioners are ahead of the academics in detecting deception. It is time for us to catch up.

## Acknowledgments

This research was conducted with the support of the National Science Foundation (SBE0725685).

## References

- Aamodt, M. G., & Custer, H. (2006). Who can best catch a liar. *Forensic Examiner*, **15**, 6–11.
- Asch, S. E. (1956). Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychological Monographs: General and Applied*, **70**, 1–70.
- Bond, C. F., Jr., & DePaulo, B. M. (2006). Accuracy of deception judgments. *Review of Personality and Social Psychology*, **10**, 214–234.
- Bond, C. F., Jr., & DePaulo, B. M. (2008). Individual differences in judging deception: Accuracy and bias. *Psychological Bulletin*, **134**, 477–492.
- Bond, C. F., Omar, A., Urvashi, P., Lashley, B. R., Skaggs, L. M., & Kirk, C. T. (1992). Fishy-looking liars: Deception judgment expectancy violation. *Journal of Personality and Social Psychology*, **63**, 969–977.
- Brunswick, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review*, **62**, 193–217.
- DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbrick, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin*, **129**, 74–118.
- Ekman, P., & Friesen, W. V. (1969). Nonverbal leakage and clues to deception. *Psychiatry*, **32**, 88–106.
- Frank, M. G., & Feeley, T. H. (2003). To catch a liar: Challenges for research in lie detection training. *Journal of Applied Communication Research*, **31**, 58–75.
- Gigerenzer, G. (2008). *Rationality for mortals: How people cope with uncertainty*. Oxford, UK: Oxford University Press.
- Honath, F.S., Jayne, B.P., & Buckley, J.P. (1994). Differentiation of truthful and deceptive criminal suspects in behavioral analysis interviews. *Journal of Forensic Sciences*, **39**, 793–807.
- Krippendorff, K. (2004). *Content analysis*. Thousand Oaks, CA: Sage.
- Levine, T. R. (2007). *Trivia game interviews: Year 2*. Unpublished videotapes. East Lansing, MI: Michigan State University.
- Levine, T. R. (2010). A few transparent liars. In C. Salmon (Ed.), *Communication Yearbook*, **34** (pp. 40–61). Thousand Oaks, CA: Sage.
- Levine, T. R., Anders, L. N., Banas, J., Baum, K. L., Endo, K., Hu, A. D. S., et al. (2000). Norms, expectations, and deception: A norm violation model of veracity judgments. *Communication Monographs*, **67**, 123–137.

- 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.
- Levine, T. R., Kim, R. K., & Hamel, L. M. (2010). People lie for a reason: An experimental test of the principle of veracity. Manuscript submitted for publication.
- 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., Park, H. S., & McCornack, S. A. (1999). Accuracy in detecting truths and lies: Documenting the "veracity effect." *Communication Monographs*, **66**, 125–144.
- Milgram, S. (1969). *Obedience to authority*. New York: Harper.
- Park, H. S., Levine, T. R., McCornack, S. A., Morrison, K., & Ferrara, M. (2002). How people really detect lies. *Communication Monographs*, **69**, 144–157.
- Ross, L., & Nisbett, R. (1991). *The person and the situation*. New York: McGraw Hill.
- Simon, H. A. (1990). Invariants of human behavior. *Annual Review of Psychology*, **41**, 1–19.
- Sporer, S. L., & Schwandt, B. (2006). Paraverbal indicators of deception: A meta-analytic synthesis. *Applied Cognitive Psychology*, **20**, 421–446.
- Sporer, S. L., & Schwandt, B. (2007). Moderators of nonverbal indicators of deception: A meta-analytic synthesis. *Psychology, Public Policy, and Law*, **13**, 1–34.
- Vrij, A. (2000). *Detecting lies and deceit: The psychology of lying and the implications for professional practice*. New York: Wiley.

## 背景内容信息能够改善测谎准确率

### 【摘要：】

以往的研究表明，人们不太善于辨别真实和谎言。然而，如果背景知识被用于判断情境信息的真实性，就能够出现较高的准确率。背景内容的有用性在一系列对学生（N=26，45，51，25，127）和专家（N=66）的试验中得到体现。在整个研究中，有背景内容的一组平均精确度为75%，而控制组则只有57%。这些结果表明在有意义的情境中进行情境判断的重要性，并对欺骗理论有重要意义。

## Résumé

La recherche a montré que les gens n'arrivent à distinguer la vérité du mensonge que dans une proportion à peine supérieure à celle obtenue par le simple effet du hasard. Cependant, des niveaux d'exactitude plus élevés sont possibles lorsque les connaissances contextuelles sont utilisées pour évaluer la véracité du contenu d'un message situé. L'utilité du contenu en contexte est démontrée dans une série d'expériences menées auprès d'étudiants ( $N = 26, 45, 51, 25, 127$ ) et d'experts ( $N = 66$ ). Dans l'ensemble des études, l'exactitude moyenne était de 75 % dans les groupes de contenu en contexte, contre 57 % dans les groupes de contrôle. Ces résultats démontrent l'importance de situer les évaluateurs dans un contexte ayant un sens. Ils ont des conséquences importantes pour la théorie du mensonge.

*Mots clés :* mensonges, mentir, détection du mensonge

Forschungsergebnisse zeigen, dass Menschen nur weniger besser als zufällig Wahrheit von Lügen unterscheiden können. Höhere Trefferquoten werden allerdings möglich, wenn kontextuelles Wissen genutzt wird, um die Wahrhaftigkeit von Botschaftsinhalten einzuschätzen. Die Nützlichkeit der Kontextualisierung von Inhalten wurde in einer Serie von Experimenten mit Studierenden (N = 26, 45, 51, 25, 127) und Experten (N = 66) gezeigt. Über die Studien hinweg lag die durchschnittliche Genauigkeit bei 75% in den Kontextualisierungsgruppen im Vergleich zu 57% in den Kontrollgruppen. Die Ergebnisse zeigen die Bedeutung der Platzierung von Entscheidern in einen bedeutungsvollen Kontext und haben wichtige Implikationen für die Täuschungstheorie.

Schlüsselbegriffe: Lügen, lügen, Lügendetektion

## 요약

과거의 연구는 사람들은 거짓말로부터 진실을 구별하는 기회가 별로 크지 않다는 것을 보여주고 있다. 그러나 높은 정도의 정확도 비율은 문맥적 지식이 상황적 메시지 내용의 진실성을 판단하는데 사용될 경우에는 높은 것으로 나타난다. 문맥에서의 내용의 사용성은 학생들과 전문가들을 대상으로 한 일련의 조사들에서 나타났다. 교차연구에서, 평균 정확성은 통제된 집단에서는 57%로 나타난 반면 문맥집단내에서의 내용에서는 75%로 나타났다. 이러한 결과들은 의미있는 문맥내에서의 상황적 판단의 중요성을 보여주는 것이며 기만이론을 위한 주요한 함의를 가지고 있는 것으로 볼 수 있다.



## Resumen

Las investigaciones anteriores han mostrado que la gente es solo un poco mejor que la casualidad para distinguir las verdades de las mentiras. Los índices de mayor precisión, sin embargo, son posibles cuando el conocimiento contextual es usado para juzgar la veracidad del contenido de los mensajes situados. La utilidad del contenido en contexto fue demostrado en una serie de experimentos con estudiantes ( $N = 26, 45, 51, 25, 127$ ) y expertos ( $N = 66$ ). A través de los estudios, el promedio de la exactitud fue del 75% en el contenido de grupos de contexto comparado con el 57% de los de control. Los resultados demuestran la importancia de situar a los jueces dentro de un contexto significativo y tiene importantes implicancias para la teoría de la decepción.

*Palabras claves:* mentiras, mentir, detección de mentiras