HUMAN
COMMUNICATION RESEARCH

Human Communication Research ISSN 0360-3989

ORIGINAL ARTICLE

Increasing Deception Detection Accuracy with Strategic Questioning

Timothy R. Levine, Allison Shaw, & Hillary C. Shulman

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

One explanation for the finding of slightly above-chance accuracy in detecting deception experiments is limited variance in sender transparency. The current study sought to increase accuracy by increasing variance in sender transparency with strategic interrogative questioning. Participants (total N = *128) observed cheaters and noncheaters who were questioned with either indirect background questions or strategic questioning. Accuracy was significantly below chance (44%) in the background questions condition and substantially above chance (68%) in the strategic interrogative questioning condition. The results suggest that transparency can be increased by strategic question asking and that accuracy rates well above chance are possible even for untrained judges exposed to only brief communications.*

doi:10.1111/j.1468-2958.2010.01374.x

It is both commonly accepted and well documented that people are only slightly better than chance at distinguishing truths from lies in deception detection experiments. Bond and DePaulo's (2006) meta-analysis found that average accuracy in deception detection experiments is only 54%, where 50% could be obtained by chance. This finding is remarkably stable, with 90% of published studies reporting results within ±10% of the across-study mean. Furthermore, meta-analysis shows that nonverbal training improves accuracy, on average, by a mere 4% (Frank & Feeley, 2003), and there is very little variance attributable to individual differences in judge ability (Bond & DePaulo, 2008) or judge professional experience (Aamodt & Custer, 2006; Bond & DePaulo, 2006). In turn, the lack of variance in judge ability results in small standard errors in accuracy scores and consistent findings across studies (Levine, 2010; for a contrasting perspective, see O'Sullivan, 2008).

Theoretically, above-chance accuracy is both expected and explained in terms of judge ability to spot sender leakage and deception cues (Ekman & Friesen, 1969; Zuckerman, DePaulo, & Rosenthal, 1981).[1] The prevailing presumption guiding the past 40 years of deception research is that lying is associated with a variety of internal psychological states including arousal, emotions such as fear or guilt, and cognitive

Corresponding author: Timothy R. Levine; e-mail: levinet@msu.edu
This research was conducted with the support of the National Science Foundation (SBE0725685). The authors thank J. Pete Blair and Judee K. Burgoon for their advice in designing the interview questions.

load (Zuckerman et al., 1981), and that these psychological states are causally associated with a variety of overt behavioral displays. Although liars are thought to try to suppress these deception-linked behaviors (Ekman & Friesen, 1969) and strategically alter their behaviors to appear honest (Burgoon & Buller, 1996), behavioral cues to deception nevertheless leak out. This usually occurs through nonverbal channels, which are more difficult for liars to actively monitor and control (Ekman & Friesen, 1969). Thus, if leakage and deception cue theories are correct, individuals skilled at spotting leakage should be able to adeptly, albeit imperfectly, distinguish truths from lies at least under conditions conducive to leakage such as those where the stakes are consequential (O'Sullivan, 2008).

Levine (2010) recently offered a different perspective. He argues that if theory based on leakage and deception cues had broad utility, then the accuracy ceiling would be higher, nonverbal training would be more fruitful, and there would be larger individual differences in message judges' ability to detect deception than is evident in meta-analyses. In addition, little evidence of specific and consistent nonverbal leakage cues have been obtained from decades of nonverbal deception cue studies (DePaulo et al., 2003; Sporer & Schwandt, 2006, 2007). Simply put, theory based on nonverbal leakage and deception cues appears implausible, given the cumulative findings of seven recent meta-analyses (see Aamodt & Custer, 2006; Bond & DePaulo, 2006, 2008; DePaulo et al., 2003; Frank & Feeley, 2003; Sporer & Schwandt, 2006, 2007).[2] Levine (2010) explains stable, slightly above-chance accuracy in terms of a few transparent liars. Transparency is a characteristic of senders. Transparent senders are those who are correctly seen as honest when telling the truth but give themselves away when they lie. Transparency can be thought of as being either a relatively stable individual difference (i.e., some people are typically more transparent than others) or variable across situations (i.e., the same person can be transparent in one situation but not another).

Transparency can be distinguished from judge ability, demeanor, and leakage. Variance in deception detection accuracy can stem from sender transparency, judge ability, or both. The term transparency is used to distinguish the variance in detection accuracy that is attributable to senders from that resulting from judge ability. The difference between transparency and demeanor is that demeanor is *independent* of honesty-deception, whereas transparency is *conditional* on honesty-deception. People with an honest demeanor tend to be believed regardless of their veracity. Thus, demeanor impacts perceived honesty, whereas sender transparency and judge ability impact accuracy. Finally, although leakage and deception cues may be sources of transparency, transparency can result from factors other than deception-linked psychological states such as emotions, arousal, and cognitive effort such as message content or context. Furthermore, although leakage is something that liars do because they are lying, exculpatory message content by honest people can make them transparent.

Levine (2010) speculates that if leakage and deception cues were typically limited to a relatively few transparent liars, then results from recent meta-analyses cohere

and become understandable. If, for example, approximately 10% of people were leaky, then judges could correctly and reliably distinguish truths from lies in this subpopulation of poor liars. In contrast, judges would perform at chance level for the remaining nonleaky 90% of the population. This would produce stable accuracy rates near 55%. That is, a 50% chance-level hit rate for 90% of the people yields 45%, and near 100% accuracy for the 10% leaky liars adds 10% to produce the stable 55% accuracy observed in meta-analyses. This view explains both why accuracy is consistently above chance and why the accuracy ceiling is so low. This also explains why there are few individual differences in detection ability, why training fails to improve accuracy by much, and why cue meta-analyses fail to see meaningful levels of leakage (Levine, 2010). According to this view, leakage and deception cues exist, but they are atypical and are not true of most liars most of the time.

The empirical stability of previous accuracy findings coupled with the disconfirmation of widespread, pervasive deception leakage might lead to a pessimistic view of human lie detection. Such a pessimistic view, however, may be premature. If Levine (2010) is correct that the low accuracy ceiling is a function of a few transparent liars under the conditions typical in the literature, then it is reasonable to question if variance in transparency might be strategically increased by altering the conditions under which lies are told and maintained. That is, it might be possible to make senders more transparent. Doing this involves both getting liars to give themselves away and providing honest senders opportunity to convey their honesty. The current study provides a test of the proposition that strategic questioning of a potential liar might enhance accuracy by increasing the proportion of transparent senders.

Research predictions

In the experiment reported here, judges were shown videotaped interviews in which students were asked about their performance on a task they had just completed. Half of the students had cheated on the task and the other half had not, but all denied cheating. Judges watched video clips either of responses to task-related background questions or of direct strategic questioning regarding cheating and honesty. The research questions of primary interest are how question type impacts deception detection accuracy and individual differences in judge ability, sender believability (demeanor), and transparency (detectability). The latter issue is based on the premise that accuracy is more a function of individual differences in message senders than individual differences in message judges (Bond & DePaulo, 2008; Levine, 2010).

Specific and formal hypotheses were avoided because several different predictions are possible. For judges exposed only to the indirect background questions, verbal content is likely of little use because the questions and answers do not bear upon the issue of cheating. Consideration of this alone would lead to a prediction of chance-level accuracy in identifying cheaters. However, from a traditional deception cue perspective, cheaters have guilty knowledge that the noncheater lacks. If cheaters felt guilty, anxious, or both, and if this guilt and anxiety were leaked, then above-chance

accuracy would be anticipated to the extent that judges recognized the guilty behavior as diagnostic. Alternatively, people who know they have an honest demeanor may be more likely to cheat and lie as personal experience tells them that they are unlikely to be caught (Serota, Levine, & Boster, 2010). Such self-selection could lead to below-chance accuracy because more believable people are more likely to cheat, and less honest-appearing people actively avoid situations requiring deception. Thus, plausible arguments can be made for above-chance, below-chance, and at-chance accuracy.

A wide range of predictions are also possible for those viewing the direct and strategic interrogative questions and answers. Clearly, previous research leads to a prediction of slightly above-chance accuracy. To the extent that previous findings generalize to more interrogative situations regardless of questioning strategy, slightly above-chance accuracy is just what ought to be expected. However, it is plausible that direct and pointed questioning may increase variance in transparency. For example, Vrij et al. (2009) show that unanticipated questions can yield diagnostically useful information and Hartwig, Granhag, Stromwall, and Kronkvist (2006) show that the strategic use of evidence (SUE) during questioning can yield accuracy rates of over 80%. To the extent that it is relatively more difficult to credibly maintain a lie under strategic, direct, sustained, and unanticipated questioning, the proportion of transparent liars is likely increased, and thus improved accuracy is expected.

Then again, the behavioral adaptation explanation (BAE; Buller, Stiff, & Burgoon, 1996) and information deception theory (IDT; Buller & Burgoon, 1996) predict just the opposite. According to the BAE and IDT, suspicious questions lead suspected liars to strategically adapt their behavior so as to appear more honest. Judges then interpret the honest-appearing behaviors as indicative of honesty, thereby increasing truth-bias and reducing accuracy under direct questioning. So once again, plausible arguments can be made for a variety of outcomes.

Given the above discussion, the current experiment seeks to answer two key questions:

RQ1: Does the type of questioning affect deception detection accuracy?

RQ2: What is the nature of sender and judge variation in demeanor/transparency under background and direct questioning?

Method

Participants

The participants in this experiment were $N = 128$ undergraduate students at a large university in the Midwestern United States. Of these, 38 (29.7%) were male and 90 (70.3%) were female. The average age was 20.0 years (range, 18–31, $SD = 2.0$). All participants received research credit for class in exchange for their participation, and the data collection was internal review board approved.

Design

The research design was a 2 × 2 mixed experiment with interviewee guilt (cheater or noncheater) as a repeated factor crossed with question type (background, nonaccusatory or direct and strategic) as an independent groups factor. There were $n = 64$ participants in each of the questioning conditions (19 males and 45 females in both) who viewed and judged 44 interviews (22 noncheaters and 22 cheaters) each.

Stimulus materials

The 44 videotaped truths and lies were selected from a data set of 111 tapes created by T. Levine at Michigan State University in 2007 and funded by the National Science Foundation. Briefly, undergraduate students played a trivia game for a cash prize with a partner as part of a study supposedly on teamwork. All subjects were provided with the opportunity to cheat and were encouraged to do so by their partner, a research confederate. After the trivia game was completed, participants were interviewed, and the interview was videotaped in high definition. Participants were told that the interview was about the role of teamwork. Each interview involved the same 10 questions asked by the same female interviewer. The first three questions asked were about prior experience with teamwork and playing trivia games. These served as the background questions in the current experiment. The next four questions asked were about the interviewee's performance in the game, but did not directly ask about cheating. These were not included in the current experiment. The last three questions asked were if cheating occurred, if the participant answered honestly, and about what the person's partner would say when he or she was interviewed. These last three questions constituted the direct and strategic questioning in the current experiment. Because all subjects decided for themselves whether or not to cheat, and if they cheated, whether or not to lie, the lies were completely unsanctioned. Because the lies and the accusations of lying were about students cheating in federally funded research at a university, participants experienced the situation as both highly involving and one of high stakes.

Every interviewee ($n = 22$) who both cheated and lied was included in the current study. Because individual differences in performance were of concern, no efforts were made to screen out poor liars. Those who cheated and confessed under questioning ($n = 7$), however, were excluded. Thus, there were 29 cheaters including those who confessed. Only lying cheaters ($n = 22$), however, were included. For each liar, a matching truthful interview was selected from the database such that the deceptive and truthful interviewees were matched on sex, race-ethnicity, and approximate physical appearance-attractiveness. In all cases, the most proximate (in terms of subject number) matching truth-teller was chosen without consideration of performance. Matching was done based on a still-frame screen shot of the interview.

The interview tapes were edited so that only the first three background questions or the last three direct questions were included. The identity and order of the 44 honest and deceptive interviewees were held constant across questioning conditions. The same interviewer interviewed all 44 interviewees, asking the same scripted questions.

The three background questions included: How much experience have you had with doing teamwork activities, can you describe your past experiences? Have you played Trivial Pursuit or games like this one in the past? Why do you (or why don't you) play trivia games often? (as a follow-up to the previous answer). The two direct questions and one strategic question were as follows: Did any cheating occur when the experimenter left the room? Are you telling me the truth? What will your partner say when I ask her the same question?

The average duration of background question and answers was approximately 1 minute, and the average duration of the direct questioning was approximately 20 seconds. The duration of the interview, however, did not impact detection accuracy beyond the two inductions, $F\ (1, 83) = 0.01$, $p = ns$, $\eta^2 = .00$, and the inclusion of time as a covariate in the analyses did not alter the results. The selected interviews were burned onto a DVD for use in the experiment along with instructions, and background information on the trivia game including the potential for cheating.

Procedures

After gaining informed consent and being provided with experimental instructions including information about the trivia game and the interview, participants watched a series of 44 videotaped interviews on a large-screen projection system. All subjects were told:

> You will view several video clips of interviews with people who had just finished playing a trivia game for a cash prize. All were given the opportunity to cheat. Some of them cheated, some of them did not.

Participants, based on random assignment, watched either the background questions and answers or the direct questions and answers. Participants in the background condition were further instructed as follows:

> You will only see the first few background questions. Even though they are not asked directly if they cheated, see if you can tell who cheated. Who is acting guilty? After each clip, decide if the person cheated or not. Mark that on your survey. Answer for every person.

Alternatively, the participants in the direct questioning condition were told the following:

> Each person is directly asked if they cheated. All deny cheating. See if you can tell who is lying. After each clip, decide if the person cheated or not. Mark that on your survey. Answer for every person.

Participants in the background condition made a forced choice cheater-or-noncheater judgment, whereas those who saw direct questioning made a lying-cheater-or-honest-noncheater judgment. Filler segments were interspersed between each interview and contained text telling the viewer to rate the previous interview. The judgments were

scored for truth-bias and accuracy. Truth-bias was simply the percent judged to be honest noncheaters and accuracy was the percent correct.

Results

Across conditions, the results closely mirrored the previous literature. Overall accuracy was 55.9% (95% CI = 53.5 to 58.3%), which was statistically greater than chance, $t(123) = 8.48$, $p < .001$. Participants were also truth-biased, judging 59.5% of interviewees as honest or noncheaters. The veracity effect (Levine, Park, & McCornack, 1999) was also evident, with truth accuracy at 65.4% and lie/cheating accuracy at 45.9%. Truth accuracy was greater than chance, $t(125) = 9.76$, $p < .001$, and lie accuracy was below chance, $t(125) = -2.36$, $p = .02$.

The first research question asked if accuracy differed depending on whether participants were exposed to background questioning or direct, strategic questioning. The relevant findings are summarized in Table 1. Somewhat surprisingly, truth-bias was nearly constant across conditions, $t(122) = 0.22$, $p = .82$. Accuracy, however, differed dramatically. Participants were substantially more accurate in the direct, strategic questioning condition (68.0%) than in the background questions condition, 44.1%, $t(122) = 20.51$, $p < .001$. One-sample t tests showed dramatically greater than chance accuracy for direct and strategic questioning, $t(60) = 21.39$, $p < .001$, and below-chance accuracy under background questions, $t(62) = -7.30$, $p < .001$. Even lie accuracy (57.5%) was greater than chance under direct and strategic questioning, $t(61) = 3.70$, $p < .001$, although the veracity effect was evident across questioning conditions. Finally, accuracy in the direct questioning conditions was also significantly greater than the 54% meta-analysis average, $t(60) = 16.64$, $p < .001$.

The second research question asked about individual differences in sender and judge ability, transparency, and demeanor. Mean ratings for individual interviewees are presented in Table 2. As can be seen, substantial individual differences are apparent and these vary systematically by questioning and veracity conditions.

Table 1 Truth-Bias and Accuracy by Questioning Condition

Condition	Truth-Bias (%)	Overall Accuracy (%)	Truth Accuracy (%)	Lie Accuracy (%)
Background				
M	59.3[a]	44.1[a]	53.4[b]	34.7[a]
SD	13.68	06.4	14.5	15.5
Direct interrogation				
M	59.8[a]	68.0[a]	77.4[a]	57.5[a]
SD	11.3	06.6	11.3	15.9

[a]A mean that is significantly different from the 50.0% chance base-rate at $p < .001$, two-tailed.
[b]$t(62) = 1.85$, $p = .069$.

Table 2 Individual Differences in Transparency and Demeanor

Cheaters–Liars			Honest Noncheaters		
Interview	Direct (%)	Background (%)	Interview	Direct (%)	Background (%)
23	95.3	50.0	20	100.0	56.3
11	92.1	53.1	5	96.9	82.8
40	85.7	29.7	6	96.9	73.4
38	84.4	09.4	39	96.9	26.6
4	79.7	32.8	10	95.3	82.8
43	79.7	42.2	17	90.6	40.6
26	78.1	18.8	13	89.1	43.8
34	76.6	42.2	33	87.5	68.8
19	73.4	28.1	24	85.9	85.9
28	65.6	28.1	15	84.4	60.9
36	62.5	23.4	21	84.1	46.0
8	60.6	57.8	12	81.3	40.6
7	56.3	57.8	29	76.6	43.8
37	56.3	39.1	27	75.0	64.1
2	37.5	45.3	35	75.0	39.6
22	34.4	23.4	25	70.3	26.6
44	34.4	14.1	32	70.3	23.4
14	31.3	45.3	41	68.8	60.9
31	31.3	46.9	9	65.6	29.7
16	29.7	40.6	1	48.4	34.4
30	10.9	07.8	3	35.9	84.4
42	07.8	28.1	18	34.4	60.9
M	57.5	34.7		77.4	53.4
SD	26.3	14.9		18.7	20.3

Note: Table values are the percent of judges ($n = 64$) correctly identifying the interviewee as a liar–cheater or as a noncheater.

Individual differences were more pronounced and extreme under direct questioning. For the cheaters, no individual was highly transparent under indirect, background question and answer. The least credible sources were judged as noncheaters by over 40% of the judges. However, under direct and strategic questioning, a number of cheaters were accurately identified as liars by a substantial proportion of the judges. Similarly, under direct and strategic questioning, more than half of the innocent interviewees were correctly identified as honest by more than 80% of judges. Nevertheless, substantial individual differences in transparency were evident within both conditions. A majority of judges consistently misperceived a minority of interviewees even under direct and strategic questioning. By contrast, there was substantially less variance in judges than senders. Table 3 shows the distributions and variances of both senders and judges in each condition. It can also be seen that questioning impacts sender variance much more than judge variance.

Table 3 Distribution of Judge Ability and Sender Transparency

	Judge Ability		Sender Transparency	
Accuracy	Background Judges (%)	Direct Judges (%)	Background Senders (%)	Direct Senders (%)
0–10%			4.5	2.3
11–20%			4.5	2.3
21–30%	3.2		22.7	2.3
31–40%	34.9		9.1	15.9
41–50%	49.2	1.6	27.3	2.3
51–60%	11.1	13.1	9.1	4.5
61–70%	1.6	54.1	11.4	15.9
71–80%		31.1	2.3	18.2
81–90%			9.1	18.2
91–100%				18.2
Variance	40.0	43.6	400.0	615.0
M	44.1	68.3	44.1	68.3

Discussion

The current experiment investigated if and how deception detection accuracy, judge ability, and sender transparency are impacted by different questioning styles. Previous research has consistently reported slightly better than chance accuracy, suggesting little, but not zero, transparency on the part of liars. It was reasoned that strategic direct questioning might increase the proportion of liars who were transparent, thereby increasing accuracy.

There is a growing consensus that questioning style meaningfully impacts detection accuracy (Hartwig et al., 2006; Vrij et al., 2009; Vrij, Mann, Kristen, & Fisher, 2007). However, the questioning approach used in the current study is conceptually different from previous studies. The inspiration for the current strategic questioning was the Reid technique. The Reid technique (Inbau, Reid, Buckley, & Jayne, 2001) is the most widely known and used, but it is substantially more elaborate and time consuming than the current questioning. The Reid technique is designed primarily to induce confession by building themes, but, consistent with the current thinking, it also involves specific questions thought to increase transparency. More recently, the SUE technique (Granhag, Stromwall, & Hartwig, 2007; Hartwig et al., 2006) appears especially promising, but it is limited to situations where the lie detector has withheld evidence that can be compared with an interviewee's statements. In the current approach, it is merely suggested that evidence may be forthcoming. In contrast to Vrij et al. (2007), the current direct questions do not fit neatly within either information gathering or accusatory styles. The current direct questioning involved brief open- and closed-ended questions that were not directly accusatory, but the final question was strategically structured to increase transparency.

The data were consistent with the prediction that direct and strategic questioning can increase accuracy by producing elevated interviewee transparency. Overall, accuracy approached 70% ($M = 68\%$, 95% CI = 66.4 to 69.7%) in the direct questioning condition. This value substantially exceeds both chance levels and levels typical of the literature. In fact, the lower bound of the current confidence interval was 12% greater than the stable across-study mean obtained from meta-analyses. Furthermore, the current findings were obtained from a relatively large number ($k = 44$) of truthful and deceptive interviews, and although judges were still truth-biased (60% of the interviewees were rated as honest), accuracy for lies (57.5%) was nevertheless significantly above chance.

The relatively high accuracy observed in the direct questioning condition is especially impressive given: (a) the brief duration of the questions and answers (the majority were under 20 seconds), (b) that the questioning was by a young graduate student rather than by a professional interrogator, (c) that the question content was held constant across interviewees, and (d) that the judges were untrained undergraduate students. By implication, extended interviews by experienced interrogators free to pursue suspicious answers might yield even higher accuracy rates than current values. Thus, past research showing poor performance from experts may not generalize to all interrogative situations. Usually, of course, experts in deception detection studies do not get to ask their own questions (for a notable exception, see Hartwig, Granhag, Stromwall, & Vrij, 2004).

Furthermore, the 68% accuracy underestimates the actual effectiveness of the direct questioning because direct questioning also yielded seven confessions that were not used in the current study. Confessions were only obtained in response to direct questioning, and no false confessions were obtained. Thus, on average, direction questioning led to the correct identification of 12.7 of 22 lying cheaters and 7 of 7 confessing cheaters yielding an adjusted detection rate of 67.8% for cheaters. If solicited confessions are included, then overall accuracy in the direct questioning condition improved to 72%.

The improvement in accuracy from background questioning to direct questioning stems from increased variance in sender transparency, which is clearly depicted in Tables 1 and 2, and especially Table 3. The variance in hit rate was four times larger for directly questioned cheaters ($V = 691.7$) than for cheaters in the background condition ($V = 222.0$). Comparatively, the variance for total judge accuracy in the two conditions was $V = 43.6$ and 41.0. No cheater was correctly identified as such by more than 60% of the judges under background questioning, and only 4 of 22 (18%) interviewees were correctly identified by at least 50% of the judges. In contrast, under direct questioning, 14 of 22 (64%) cheaters were correctly identified by at least 50% of the judges, and of those, 8 were identified correctly by 75% or more of the judges. For honest, noncheating interviewees, 6 were correctly identified by more than 90% of the judges under direct questioning. No interviewee was judged correctly by more than 90% of the judges under background questioning. Clearly,

direct questioning made a larger proportion of the interviewees more transparent than either background questioning or than is typical in previous studies.

Consistent with the distinction between transparency and leakage, it is noteworthy that much of the improvement in accuracy was attributable to correctly identified truth-tellers. Recall that leakage and deception cues are something that liars do, while transparency applies to both truth-tellers and liars. Under direct questioning, 6 of 22 (27.2%) noncheaters were correctly identified by 90% or more of the judges and a majority (54.5%) of noncheaters were correctly identified by at least 80% of the judges. In contrast, truth accuracy rates were lower under background questioning with no noncheater being correctly identified by more than 86% of the judges. This finding is not attributable to truth-bias because truth-bias did not differ across conditions. Furthermore, it is unlikely that the improvement was the result of a mere absence of leakage. Instead, the strategic questioning apparently aided honest sources in providing exculpatory answers in addition to revealing the lies of some of the liars.

An intriguing question is why judges were significantly less accurate than chance in the background question condition. If one rejects the idea of nonverbal leakage of guilty knowledge, then the lack of statistically significant accuracy in this condition may not be especially surprising given that message content was of little use. But, this does not explain why cheaters were rated as noncheaters with greater frequency than actual noncheaters. Two potential explanations are considered here.

The first explanation is a variation on the BAE (Buller et al., 1996). This account holds that cheaters likely knew that they might be under suspicion while noncheaters had no reason to believe that they were a suspected cheater. Cheaters, then, might have made a conscious effort to appear like a noncheater, and hence were actually more likely to be seen as noncheaters than the actual noncheaters who were not consciously aware of the need to act honest. The net result was below-chance accuracy. However, if the BAE were valid, it should have been in operation in an even stronger form under direct questioning because perceived suspicion is a necessary antecedent for the BAE (Buller et al., 1996). The direct questioning certainly communicated that the interviewees were under suspicion, yet they were not perceived as more honest under direct questioning.

A more likely explanation is a self-selection hypothesis. Although it is commonly believed that most people lie daily (cf. DePaulo, Kashy, Kirkendol, Wyer, & Epstein, 1996), recent research suggests substantial individual differences in the proclivity to lie and that most lies are told by relatively few prolific liars (Serota et al., 2010). Although the average number of lies told per day may be greater than 1.00, the distribution of lying is highly skewed with most people communicating honestly most of the time.

Clearly, there is substantial variance in sender ability to lie well (i.e., transparency and demeanor; Bond & DePaulo, 2008; Levine, 2010; current findings). Furthermore, there likely exists a substantial correlation between the ability to lie well and the willingness to lie because practice tends to improve skill (obviously, prolific liars get much more practice) and because people often avoid that which they are not good at

(Levine, 2010; Serota et al., 2010). This is especially true for cheating and lying when the consequences of detection can be serious. Furthermore although judges often do not get good feedback on their detection ability, liars usually know when their lies are discovered and consequently, people likely know if they are a good liar or not. All this suggests that people with a more honest demeanor were more likely to cheat than people with a less honest demeanor, explaining below-chance accuracy under background questioning.

If this self-selection reasoning is accurate, it makes the detection rates under direct questioning more impressive still. This would mean that direct questioning yielded substantially better than average accuracy on a sample of better than average liars.

A second intriguing question is if combining background and direct questioning would yield higher or lower accuracy than viewing the answers to the direct questions only. Having both types of questions would allow for a comparison of individual honesty baselines. Early research by Miller (e.g., Brandt, Miller, & Hocking, 1980) suggested that exposure to baseline truthful answers improved accuracy somewhat. A rival hypothesis, however, is that the inclusion of additional nondiagnostic, nontransparent behaviors might be distracting, diluting, or both. Judges might form initial inaccurate impressions based on initial baseline information and this might dilute or preempt the diagnostic value of the answers to the direct questions. Future research might include a background plus direct questions condition.

The current findings have important implications for deception theory. The finding of below-chance accuracy in the background question condition meshes poorly with the traditional emotion-based leakage and deception cue perspective. Although people in this condition were not lying per se, they did have guilty knowledge in a high-stakes situation. Furthermore, although it was clear that many liars did leak the truth under direct questioning, leakage was prompted by specific questions rather than naturally occurring because of internal psychological states associated with mere deception or mere guilt. Finally, it would be inaccurate to assume that leakage was primarily nonverbal. Watching the especially transparent truth-tellers and liars revealed that, at least in some cases, the nature of leakage was in verbal content, not just nonverbal presentation. Thus, the current findings offer little support for leakage-based deception theory.

Although the current results provide a dramatic example of the utility of viewing deception from an interactive perspective, the results contradict IDT logic in important ways. IDT, for example, predicts that liars adapt to skeptical interrogative questions by adjusting their behaviors to appear more honest (i.e., the BAE; Buller & Burgoon, 1996; Buller et al., 1996). In the current study, directly interrogative questioning yielded higher accuracy, an outcome directly opposite from BAE and IDT predictions. Also, IDT is a sender–receiver interaction model predicting that accuracy is a function of the relative skills of each. The current findings are more in line with a sender variance model. Judge skill appears less important than sender skill. Most judges pick up on highly transparent senders, and most judges miss those senders where sender demeanor does not match sender honesty.

A second issue with IDT in relation to the current findings is that most IDT-based detection studies have senders and judges interact where each sender is judged by a single judge, each judge evaluates a single sender, and sender and judge are not crossed in the research design with each sender–judge dyad independent of other dyads. Such designs fully confound sender effects, judge effects, and sender–judge interaction effects. That is, sender and judge variance cannot be meaningfully partitioned in most IDT research. As a consequence, even though IDT champions an interactive view of deception, IDT studies may ironically be ill suited to identifying the salient aspects of interaction that are important in deception detection (e.g., partitioning transparency from judge ability or demeanor from truth-bias).

In the current design, senders and judges were crossed with all judges viewing all senders. This allowed for the examination of variance in senders across judges and the variance in judges across senders. However, the current design could not fully partition demeanor and transparency because senders were either cheaters or noncheaters. It is clear that transparency differed across conditions, whereas mean demeanor was approximately constant as reflected in accuracy and truth-bias scores, respectively. However, at the level of the individual sender, the two sources of variance cannot be fully disentangled. If the goal is to understand deceptive interaction, future research will need to fully disentangle judge truth-bias, judge ability, sender demeanor, and sender transparency. This will require not only crossing senders and receivers but also crossing both with veracity in a three-way design.

Although the current research investigated the impact of questioning on deception detection, the current results are not informative about the probing effect (Levine & McCornack, 2001). The probing effect refers to the finding that questioned senders are more likely to be believed than senders who are not questioned regardless of veracity. The current study, in contrast, compared two types of questioning but did not include a no-question control.

The current findings also do not mean that all forms of direct or strategic questioning are equally diagnostic. Instead, the wording of the particular questions is crucial. In the current interviews, the critical last question asked the interviewee what his or her partner would say in response to the cheating question. Noncheaters had good reason to believe that their partners would corroborate their innocence relative to lying cheaters and thus found the question easier to answer. In a previous data set involving the cheating paradigm with a different question set, the final question asked interviewees why they should be believed. That question tended to catch honest noncheaters off guard, and detection studies using those tapes failed to find the improved accuracy (e.g., Levine, Kim, Park, & Hughes, 2006; Levine, Kim, & Blair, 2010). Thus, substantially improved accuracy from direct questioning likely requires identifying specific questions that are diagnostic and create transparency. There is good reason to believe that not all direct questions are equally useful. Furthermore, useful questions are likely context dependent. The current questioning might backfire if cheating pairs had the opportunity and foresight to coordinate their answers.

A necessary confound in the current research design involves the nature of the deception in the two questioning conditions. The direct question required cheaters to lie outright while deception in the background question condition only required omission. Thus, the comparison of accuracy under direct versus background questioning is confounded with the detection of bald-faced falsehood versus deceptive omission. As a consequence, on the surface, the inclusion of the background question condition may seem odd and a less than ideal point of comparison.

There are, however, several reasons why the inclusion of the background question condition has utility even though it necessitated confounding questions and answers. First, accuracy in both the direct and background conditions can be compared with the 50% chance rate and to the 54% mean from meta-analyses so both conditions can stand alone in addition to in comparison with one another. This allows both conditions to be interesting in their own right, and conclusions need not rest solely on a direct comparison between the experimental groups. Second, as noted in the rationale, leakage-based theories might predict that the cheaters would exhibit behavior indicative of guilt regardless of the type of questions asked. The inclusion of the background questions rules out this alternative. Third, the background condition provided a demeanor base-rate to assess the perceived honesty and variability in demeanor of the interviewee under something other than direct questioning. Fourth, Serota et al. (2010) argue that people who are likely to lie probably have more honest demeanors. The inclusion of the base-rate condition and the findings of below-chance accuracy provide data consistent with Serota et al.'s arguments. Fifth, research finds that omission is the most prevalent form of deception (Levine et al., 2002) yet the deception detection literature tends to compare outright lies with unproblematic truth (cf. McCornack, 1997). Outside the lab, most deception likely takes place when the people being duped simply do not ask and the person doing the deception simply does not tell. Given that we believe that mere omission is a very prevalent and realistic deception scenario, the background questioning is likely more typical of outside the lab deception detection than most deception research. That is, the background condition is an ecologically interesting and understudied case. Finally, as noted previously, the results in the background condition are provocative. Significantly below-chance accuracy is an unusual finding with important theoretical implications. The fact that the people in the background question condition were not outright lying per se might explain null findings, but it does not explain below-chance accuracy.

In conclusion, the current experiment found evidence that direct interrogative questioning of a potential liar is associated with detection accuracy rates substantially higher than is typical of the literature. Furthermore, the data were consistent with the idea that strategic questioning increased accuracy by increasing sender transparency. Specifically, some liars (but not all) failed to maintain credibility in the face of direct and unanticipated questioning. Strategic questioning also enhanced the appearance of honesty among most honest interviewees. Future research might investigate specific direct questioning strategies to assess the generality of the current results.

Notes

1 There are at least two uses of the term ''leakage'' in the deception literature. In the original, classic usage, Ekman and Friesen (1969) defined leakage as behaviors that inadvertently reveal the content hidden by a lie, while deception cues are those behaviors that signal lying, but not what is hidden by the lie. Over time, however, the distinction has faded somewhat and the two classes of behaviors are often used interchangeably. That is, leakage is often used as a generic label for any clue to deceit that leaks out inadvertently during the telling of a lie, which results from the act of lying, and that correctly signals deceit. Here, it is proposed that behaviors associated with guilt and anxiety stemming from a transgression other than lying (cheating) might be expected based on the same logic as deception cues.

2 There is not universal agreement on the interpretation of findings from either deception cue or deception detection studies. For example, O'Sullivan (2008) asserts a view at odds with the interpretation provided in this article. She cites evidence that (a) there are important individual differences in deception detection ability and (b) high accuracy is typically obtained with nonstudent samples in conjunction with high-stakes lies. The current perspective is that the literature is large and individual studies can be cited for a wide variety of different interpretations including those of O'Sullivan, but the preponderance of evidence (including all high-quality meta-analyses) is more consistent with the current perspective.

References

Aamodt, M. G., & Custer, H. (2006). Who can best catch a liar. *Forensic Examiner*, **15**, 6–11.

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.

Brandt, D. R., Miller, G. R., & Hocking, J. E. (1980). The truth-deception attribution: Effects of familiarity on the ability of observers to detect deception. *Human Communication Research*, **6**, 99–110.

Buller, D. B., & Burgoon, J. K.(1996). Interpersonal deception theory. *Communication Theory*, **6**, 203–242.

Buller, D. B., Stiff, J. B., & Burgoon, J. K. (1996). Behavioral adaptation in deceptive transactions fact or fiction: Reply to Levine and McCornack. *Human Communication Research*, **22**, 589–603.

DePaulo, B. M., Kashy, D. A., Kirkendol, S. E., Wyer, M. M., & Epstein, J. A. (1996). Lying in everyday life. *Journal of Personality and Social Psychology*, **70**, 979–995.

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.

Granhag, P. A., Stromwal, L. A., & Hartwig, M. (2007). The SUE technique: The way to interview to detect deception. *Forensic Update*, **88**, 25–29.

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.

Hartwig, M., Granhag, P. A., Stromwall, L. A., & Vrij, A. (2004). Police officiers' lie detection accuracy: Interrogating freely versus observing video. *Police Quarterly*, **7**, 429–456.

Inbau, F. E., Reid, J. E., Buckley, J. P., & Jayne, B. P. (2001). *Criminal interrogations and confessions*. Gaithersburg, MD: Aspen.

Levine, T. R. (2010). A few transparent liars. *Communication Yearbook*, **34**.

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., 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., Lapinski, M. K., Banas, J., Wong, N., Hu, A. D. S., Endo, K., et al. (2002). Self-construal, self and other benefit, and the generation of deceptive messages. *Journal of Intercultural Communication Research*, **31**, 29–47.

Levine, T. R., & McCornack, S. A. (2001). Behavioral adaption, confidence, and heuristic-based explanations of the probing effect. *Human Communication Research*, **27**, 471–502.

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.

O'Sullivan, M. (2008). Home runs and humbugs: Comments on Bond and DePaulo (2008). *Psychological Bulletin*, **134**, 493–497.

Serota, K. B., Levine, T. R., & Boster, F. J. (2010). The prevalence of lying in America. *Human Communication Research*, **36**, 1–24.

Sporer, S. L., & Schwandt, B. (2007). Moderators of nonverbal indicators of deception: A meta analytic synthesis. *Psychology, Public Policy, and Law*, **13**, 1–34.

Sporer, S. L., & Schwandt, B. (2006). Paraverbal indicators of deception: A meta-analytic synthesis. *Applied Cognitive Psychology*, **20**, 421–446.

Vrij, A., Leal, S., Granhag, A., Mann, S., Fisher, R. P., Hillman, J., & Sperry, K. (2009). Outsmarting the liars: The benefit of asking unanticipated questions. *Law and Human Behavior*, **33**, 159–166.

Vrij, A., Mann, S., Kristen, S., & Fisher, R. P. (2007). Cues to deception and ability to detect deception as a function of police interview style. *Law and Human Behavior*, **31**, 499–518.

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: Academic Press.