HUMAN COMMUNICATION RESEARCH

Human Communication Research ISSN 0360-3989

ORIGINAL ARTICLE

Sender Demeanor: Individual Differences in Sender Believability Have a Powerful Impact on Deception Detection Judgments

Timothy R. Levine[1], Kim B. Serota[1,2], Hillary Shulman[1], David D. Clare[1], Hee Sun Park[1], Allison S. Shaw[1], Jae Chul Shim[3], & Jung Hyon Lee[3]

1 Department of Communication, Michigan State University, East Lansing, MI 48824, USA
2 Department of Management and Marketing, Oakland University, Rochester, MI 48309, USA
3 Department of Communication, Korea University, Seoul, Republic of Korea

*Sender demeanor is an individual difference in the believability of message senders that is conceptually independent of actual honesty. Recent research suggests that sender demeanor may be the most influential source of variation in deception detection judgments. Sender demeanor was varied in five experiments (*N = *30, 113, 182, 30, and 35) to create demeanor–veracity matched and demeanor–veracity mismatched conditions. The sender demeanor induction explained as much as 98% of the variance in detection accuracy. Three additional studies (*N = *30, 113, and 104) investigated the behavioral profiles of more and less believable senders. The results document the strong impact of sender effects in deception detection and provide an explanation of the low-accuracy ceiling in the previous findings.*

doi:10.1111/j.1468-2958.2011.01407.x

> By demeanor I shall refer to that element of the individual's ceremonial behavior typically conveyed through deportment, dress, and bearing, which serves to express to those in his immediate presence that he is a person of certain desirable or undesirable qualities. In our society, the "well" or "properly" demeaned individual displays such attributes as discretion and sincerity; modesty in claims regarding self; sportsmanship; command of speech and physical movements; self-control over his emotions, his appetites, and his desires; poise under pressure; and so forth. . . . The well-demeaned individual possesses the attributes popularly associated with "character training" or "socialization," these being implanted when a neophyte of any kind is housebroken. Rightly or wrongly, others tend to use such qualities diagnostically, as evidence of what the actor is generally like.
>
> —Erving Goffman (1956, p. 489)

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

It is well known and widely accepted that most people perform poorly in lie detection experiments most of the time (Weinberger, 2010). Meta-analysis shows that people are, on average, just slightly above chance at correctly distinguishing truths from lies, and that the accuracy rates observed in most deception detection experiments fall close to the across-study average (Bond & DePaulo, 2006). Although some interesting exceptions to the only slightly better than chance conclusion have recently emerged (e.g., Blair, Levine, & Shaw, 2010; Hartwig, Granhag, Stromwall, & Kronkvist, 2006; Levine, Shaw, & Shulman, 2010), an intriguing question is why the detection ceiling is too low in the vast majority of studies.

It is proposed here that near-chance accuracy is, primarily, a function of huge variance in sender demeanor coupled with a specific research design feature that is evident in almost every previous deception detection experiment, namely, that senders are randomly assigned to veracity conditions. Demeanor, as suggested in the opening quote by Erving Goffman, is a constellation of behaviors that conveys to others a sense of the sender's character, including whether or not the sender is trustworthy and believable. To demonstrate the dramatic influence of sender variance on veracity judgments and deception detection accuracy, the experiments reported here manipulated sender demeanor. By systemically varying which senders the judges evaluate, it should be possible to produce accuracy and truth–lie bias rates that diverge substantially in both directions from the outcomes typically observed in the literature. In addition, results should replicate across judges. This research further identifies the aspects of a sender's performance that makes him or her more or less believable and cross-validates the identified demeanor cues with additional data.

Deception detection research

There exists a large literature on deception detection accuracy that spans several decades and hundreds of individual studies. In the typical deception detection experiment, some research participants are senders who either lie or tell the truth and other participants are judges who evaluate the veracity of the senders. Across hundreds of studies, accuracy in these experiments is normally distributed around 54% (Bond & DePaulo, 2006). Usually, accuracy is statistically better than chance, but accuracy seldom exceeds 65%. However, because judges are typically truth biased, honest messages are identified correctly more often than lies (''the veracity effect''; Levine, Park, & McCornack, 1999).

Theoretically, deception detection has been presumed to rest on the judge's ability to spot sender leakage and deception cues (Ekman & Friesen, 1969; Zuckerman, DePaulo, & Rosenthal, 1981). According to leakage and deception cue theories, lying is usually more arousing, cognitively effortful, and invokes more emotion than telling the truth. These deception-linked internal psychological states, in turn, are associated with specific and recognizable behavioral manifestations. Despite sender efforts to control self-presentation, clues to deception nevertheless leak out nonverbally, enabling accurate deception detection by an adept observer. In short,

the theoretical perspectives that have guided many deception detection experiments (e.g., leakage, Ekman & Friesen, 1969; four-factor theory, Zuckerman et al., 1981; interpersonal deception theory, Buller & Burgoon, 1996) specify that, at least under certain conditions, deception is detectable based on sender demeanor because the act of lying changes sender demeanor in ways that have diagnostic utility. According to these perspectives, a useful link exists between sender demeanor and sender veracity that properly perceptive judges could use to accurately detect deception.

The repeated finding that people are not much better than chance at detecting deception may appear inconsistent with the idea that lies are detectable based on behavioral observation. Therefore, the low-accuracy ceiling in the literature poses a potential problem for theory. One explanation for the low-accuracy ceiling in deception detection experiments is that, although useful nonverbal deception cues are typically present, judges in deception experiments simply look for the wrong cues (Miller & Stiff, 1993; Zuckerman et al., 1981). Efforts to improve accuracy through nonverbal training, however, have been disappointing (Frank & Feeley, 2003). Consequently, other explanations for the low-accuracy ceiling also incorporate the nature of the lies told. It is reasoned that the lies told in most deception detection experiments are innocuous falsehoods of little real consequence and therefore such lies are insufficient to produce the arousal, emotions, and cognitive effort necessary for leakage and deception cues (Frank & Feeley, 2003; O'Sullivan, Frank, Hurley, & Tiwana, 2009). According to the most contemporary version of this view, better accuracy requires both high-stakes lies and skilled judges (O'Sullivan, 2008; O'Sullivan et al., 2009).

Sender honest and dishonest demeanor

There is substantial variance in the appearance of honesty independent of actual honesty, and when this variance is coupled with random assignment of senders to veracity conditions, it provides a plausible alternative explanation for near-chance accuracy and the low-accuracy ceiling in deception detection experiments. As used in this research, and consistent with Levine (2010) and Levine et al. (2010), sender demeanor refers, specifically, to the believability of the message sender independent of actual honesty. The idea is that some people come off as sincere while others do not and, for most people, this has little or nothing to do with whether or not they are actually honest or actually lying. The observation that individual differences in sender demeanor affect deception judgments goes back, at least, to Zuckerman, DeFrank, Hall, Larrance, and Rosenthal (1979), but subsequent research has tended to focus more on individual differences in judges than in senders (e.g., detection wizards; see Bond, 2008; O'Sullivan & Ekman, 2004). An exception is a series of studies by Frank and Ekman (2004) who called the idea "truthfulness generality" and found individual differences in believability were highly stable across two situations.[1] The more general idea of individual differences in demeanor was originally explicated by Goffman (1956). The idea of demeanor presented here is consistent with that

provided by Goffman, except that in our research the focus is specifically on what makes a person come off as more or less honest in appearance. This idea differs from leakage theory, four-factor theory, and interpersonal deception theory because demeanor is presumed to be independent of actual veracity rather than a consequence of it.

Although sender demeanor is an individual difference, it is not presumed to be completely trait-like. Instead, there are likely to be situational variations and trait-by-situation interactions in demeanor.

Honest (or dishonest) demeanor is defined as a sender attribute even though its impact is assessed through the observations and judgments of message recipients. Evidence that demeanor is a sender-driven phenomenon can be obtained through observations of systematic variance in believability across senders and relative judge constancy for specific senders. That is, strong evidence for demeanor as a property of senders can be obtained by showing that different senders are seen the same way by different judges. Simply put, the logic of this approach suggests that judges are relatively interchangeable with regard to their perceptions of a specific sender, but different senders are seen differently across judges. This marks another departure from current cue-based theories, which presume that leakage and deception cues are general across senders, but that the ability to recognize cues varies across individual judges.

If senders vary substantially in demeanor and if senders are randomly assigned to truthful or lying conditions, then sender demeanor should be distributed quasi-evenly (i.e., not systematically) across the truth and lies judged. In a typical and traditional deception detection experiment, leakage and deception cues can be thought of as the deception signal, and variance in sender demeanor can be thought of as noise because it is, by our definition, unrelated to actual veracity. As the variance in sender demeanor increases relative to the amount of sender leakage, valid deception cues, sender transparency, and judge ability, the noise-to-signal ratio will also increase. Coincident with an increase in noise-to-signal ratio, the accuracy ceiling will be reduced and accuracy will be pushed down toward mere chance levels.

The consistent finding that detection accuracy is only slightly better than chance implies that the signal is not zero, but that the noise-to-signal ratio is typically high (Levine, 2010). The noise could be attributable to inexperienced, untrained, or unmotivated judges, the nature of lies (i.e., unsanctioned, low stakes), natural variance in sender demeanor, some other unknown factors, or some combination of these factors. Recent evidence suggests, however, that variance in sender demeanor is the primary factor at play, and that it swamps both the leakage signal and all other sources of noise (Bond & DePaulo, 2008; Levine et al., 2010).

Bond and DePaulo (2008) analyzed sender and judge variance in deception detection experiments. They reported that variance in sender demeanor was 10 times larger than any other single source of variance in veracity judgments and was approximately 100 times larger than the variance in judge ability. The stunning

implication of these finding is that sender demeanor may be the single most influential variable in deception detection experiments. Levine et al. (2010) replicated the finding that sender variance is 10 or more times larger than judge ability variance in an experiment using high-stakes lies. As was argued previously, the impact of variance in sender demeanor in most deception detection experiments is to disguise meaningful leakage and transparency and thereby cap accuracy. Substantial variance in sender demeanor might therefore be a sufficient and parsimonious explanation for the low-accuracy ceiling typical in the literature.

Research predictions

If sender demeanor is as important as Bond and DePaulo's (2008) and Levine et al.'s (2010) results suggest, then it should be theoretically and empirically possible to manipulate the outcome of deception detection experiments by controlling sender demeanor (Levine, 2010). In the first five experiments reported here, senders were selected from a pool of recorded interviews based on their demeanor. Four types of senders were selected using pretested results: sincere truthtellers, sincere liars, insincere truthtellers, and insincere liars. These senders were used to form four collections of senders: those whose demeanor and veracity matched (i.e., sincere truthtellers and insincere liars), demeanor and veracity mismatched (i.e., insincere truthtellers and sincere liars), sincere demeanor (i.e., sincere truthtellers and sincere liars), and insincere demeanor (i.e., insincere truthtellers and insincere liars). To the extent that outcomes can be determined by manipulation of sender demeanor, we predicted accuracy should be high for demeanor–veracity matched senders, whereas accuracy should be low for demeanor–veracity mismatched senders. We also predicted sincere demeanor senders would be believed regardless of veracity, whereas insincere demeanor senders would be judged as deceptive regardless of their honesty. The important question, however, is not if manipulating sender demeanor will affect experimental outcomes, for it surely will, but just how strong and reliable are sender demeanor effects on judge's evaluations of honesty and deceit, and to what extent are judges really interchangeable?

Creation of the demeanor induction

The materials for the deception detection task used in the first four experiments reported here were selected from a database of honest and deceptive interviews created by Levine (2007) based on an experiment originally designed by Exline, Thibaut, Hickey, and Gumpert (1970). The results of Levine et al. (2010) were used to select particular senders for use in this series of experiments. The participants in Levine (2007) were recruited for what they believed was a study of teamwork. Together with a partner, who was a research confederate, they played a trivia game for a cash prize. Early in the game, the experimenter was called out of the room, answers were left in easy reach, and the confederate suggested cheating. After the experimenter

returned, the game resumed. Following the game, participants were interviewed on camera about the role of teamwork, their performance in the game, and ultimately if cheating occurred. From the confederate's report, ground truth about whether or not each subject cheated was recorded. In the tapes, all lies were unsanctioned and the lies were of potential consequence. Cheaters cheated in federally funded research in violation of the university code of conduct, and while no penalty for cheating was ever enforced, cheaters did not know this would be the case at the time of questioning. Trivia game participants were verbally advised that the study they were participating in was federally funded and sanctioned by their university. Therefore, by cheating, the cheaters were knowingly attempting to defraud and sabotage a federally funded research project for their own financial gain. Because the experiment was materially a sting operation, no fraud actually occurred. The creation of the tapes was Institutional Review Board (IRB) approved, all participants were thoroughly debriefed, and additional consent was obtained for the use of the tapes in subsequent research. Additional IRB approval was obtained for this series of studies.

In the Levine et al. (2010) study, 44 videotaped truths and lies (22 of each) were selected from the 111 tapes in the full database. These included all 22 lying cheaters in the database. For each liar, a truthful interviewee was selected who was matched on sex, race–ethnicity, and approximate physical appearance–attractiveness. The tapes were edited so that only the questions directly asking about cheating were included. Three direct questions were asked: 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 resulting 44 interviewers were viewed by a sample of student judges ($N = 64$; see Levine et al., 2010 for details) who judged each sender as either a lying cheater or an honest noncheater. Across the 44 senders, the judges rated the senders as honest 60% of the time and achieved 68% accuracy. Accuracy was scored for individual senders, and based on the results 20 senders were selected from the 44 tested for use in these studies. Five each were selected of the most often believed truthtellers, most often believed liars, the least frequently believed truthtellers, and the least frequently believed liars.

The 20 selected interviews were used to create four experimental tapes. Each tape contained 10 interviews with five truthful senders and five cheating liars. The first tape contained 10 demeanor–veracity matched senders with five sincere truthtellers and five insincere liars. The demeanor–veracity mismatched tape contained the five insincere truthtellers and the five sincere liars. The sincere demeanor tape included the five sincere truthtellers and the five sincere liars, and the insincere demeanor tape included the five insincere truthtellers and the five insincere liars. Across the four conditions, each interview was used twice; once in the demeanor matched and mismatched pair of tapes and once in the sincere and insincere pair of tapes. These tapes were shown to different samples of judges in the first four experiments reported here. A subset of the tapes was shown to the judges in Experiment 5.

Deception detection experiments

Experiment 1

Participants

The first set of judges were $N = 30$ (21 women; age $M = 20.00$, $SD = 1.89$) undergraduate students enrolled in two small communication classes at a large university in the Midwestern United States. One class consisted entirely of freshman nonmajor honor students and the other class contained senior-level majors. Students participated in the study as an in-class activity related to course content. The data were collected during the first class session of the semester before the presentation of any relevant course content. No extra course or research credit was awarded.

Design and procedures

The research design for the first study was a fully repeated measures, four-condition experiment. After giving informed consent and being provided with experimental instructions including information about the trivia game and the interview, participants watched a series of 40 videotaped interviews (four tapes with 10 interviews each) on a large-screen projection system. Filler segments were interspersed between each interview and contained text telling the viewer to rate the previous interview. Participants made a forced choice lying-cheater or honest-noncheater judgment. The judgments were scored for truth bias and accuracy. Truth bias is measured as the percentage of all judgments in which the senders were judged to be honest noncheaters; accuracy is the percent correct.

Results and discussion

Across conditions, participants correctly identified 59.1% of the interviews. This value was statistically larger than both the 50% chance rate, $t(29) = 7.84$, $p < .001$, and the 54% Bond and DePaulo (2006) meta-analysis finding, $t(29) = 4.39$, $p < .001$. Participants also tended toward truth bias, judging 54.3% of messages as honest, but the proportion of messages judged as honest was not statistically greater than 50%, $t(29) = 1.81$, $p = .08$ (two-tailed).

In the demeanor–veracity matched condition, accuracy (78.7%) was well above both chance, $t(29) = 11.35$, $p < .001$, and meta-analysis levels, $t(29) = 9.77$, $p < .001$. In contrast, in the demeanor mismatched condition, accuracy (36.3%) was well below chance, $t(29) = -5.76$, $p < .001$. Accuracy in the two conditions was statistically and substantially different, $F(1, 29) = 132.84$, $p < .001$, $\eta^2 = .82$. Truth bias in the two conditions, however, was not statistically different, $F(1, 29) = 0.52$, $p = .47$, $\eta^2 = .00$.

In the sincere demeanor condition, truth bias (69.0% honest) was substantial and well above 50%, $t(29) = 5.19$, $p < .001$. In contrast, lie bias (35.7% honest) was observed in the insincere demeanor condition. The proportion was significantly below 50%, $t(29) = -3.78$, $p < .001$. The two conditions differed substantially in the proportion of messages judged as honest, $F(1, 29) = 49.15$, $p < .001$, $\eta^2 = .63$. The

two conditions, however, were not different in terms of accuracy, $F(1, 29) = 1.52$, $p = .23$, $\eta^2 = .05$. The findings of Experiment 1 are summarized in Table 1.

The results indicated that the demeanor induction had a strong impact on veracity judgments and detection accuracy. Furthermore, the results demonstrate that accuracy and believability can be manipulated independently through the selection and inclusion of different senders.

Central to this argument is the idea that different judges tend to see individual senders in the same way. A test of this thinking necessarily involves replicating these findings with different judges. The next experiment replicates Experiment 1 with a different student sample and an independent groups design.

Experiment 2

Participants and procedures

The second set of judges were $N = 113$ (82 women; age $M = 19.58$, $SD = 1.57$) undergraduate students enrolled in a large freshman-level class at the same large Midwestern U.S. university. The procedures were identical to Experiment 1 except that (a) the study was performed outside of class time, (b) research participants were awarded class research credit in exchange for participation, and (c) the design was carried out with independent groups rather than repeated measures. Participants were randomly assigned to one of the four conditions with cell sizes ranging from $n = 27$ to 30. Participants also made behavioral ratings that are reported later in this report.

Results and discussion

Across conditions, participants correctly identified 66.0% of the interviews as honest or deceptive. This value was statistically larger than the 50% chance rate, $t(112) = 8.62$, $p < .001$. Participants were also significantly truth biased compared with chance, judging 55.7% of messages as honest, $t(112) = 3.70$, $p < .001$.

In the demeanor–veracity matched condition, accuracy (77.7%) was well above chance, $t(29) = 11.86$, $p < .001$. In contrast, in the demeanor mismatched condition, accuracy (41.4%) was well below chance, $t(27) = -3.35$, $p = .002$. Accuracy rates in the two conditions were statistically and substantially different, $t(56) = 10.49$,

Table 1 Accuracy and Truth Bias in Experiment 1 (U.S. Students; $N = 30$)

	Accuracy		Percent Honest	
Tape	*M* (%)	*SD* (%)	*M* (%)	*SD* (%)
Demeanor–veracity matched	78.7	13.8	55.0	10.4
Demeanor–veracity mismatched	36.3	13.0	57.3	18.8
Sincere demeanor	58.3	13.4	69.0	20.0
Insincere demeanor	63.0	14.4	35.7	20.1
Across tapes	59.1	12.9	54.3	12.9

$p < .001$, $\eta^2 = .66$. Truth bias in the two conditions, however, was not statistically different, $t(56) = 0.25$, $p = .80$, $\eta^2 = .00$.

In the sincere demeanor condition, truth bias (61.1% honest) was significantly above 50%, $t(26) = 2.99$, $p = .006$. In contrast, lie bias (45.7% honest) was observed in the insincere demeanor condition. The proportion was not, however, significantly below 50%, $t(27) = -1.38$, $p = .17$. The two conditions differed substantially in the proportion of senders judged as honest, $t(53) = 3.19$, $p = .002$ (two-tailed), $\eta^2 = .16$. As expected, the two conditions were not different in terms of accuracy, $t(53) = 0.78$, $p = .44$, $\eta^2 = .01$. The findings of Experiment 2 are summarized in Table 2.

Experiments 3–5

Central to this argument is the proposition that judges' veracity assessments are more a function of variance in sender demeanor than judge ability. It follows from this proposition that manipulating the identity of the sender will have a large, predictable, and relatively constant effect across different judges. Showing the same senders to different judges is expected to yield comparatively constant results. The results of the first two experiments demonstrated this neatly with two samples of U.S. university students as the judges. However, the judges all came from the same population of university students, so it is possible that there was insufficient diversity in the judges to give this thinking a rigorous test. The next three studies involve more risky tests with two different nonstudent samples and a student sample with a different language and cultural background. If the same pattern of results observed previously extends across these notably different samples, a more convincing case for sender primacy and judge interchangeability can be made.

It has been argued recently that students and nonstudents perform differently in deception detection tasks and that findings from student samples do not generalize to older, more experienced adults (O'Sullivan, 2008; O'Sullivan et al., 2009). Although these claims are clearly inconsistent with the findings of meta-analyses (Aamodt & Custer, 2006; Bond & DePaulo, 2006, 2008), advocates of the student versus nonstudent distinction argue that the meta-analysis findings are based largely on student samples assessing low-stakes lies, and that meaningful judge effects emerge

Table 2 Accuracy and Truth Bias in Experiment 2 (U.S. Students; $N = 113$)

	Accuracy		Percent Honest	
Tape	*M* (%)	*SD* (%)	*M* (%)	*SD* (%)
Demeanor–veracity matched	77.7	12.8	58.3	10.2
Demeanor–veracity mismatched	41.4	13.5	57.5	14.6
Sincere demeanor	73.7	15.0	61.1	19.3
Insincere demeanor	70.7	13.6	45.7	16.4
Across tapes	66.0	19.8	55.7	16.3

only in the assessment of high-stakes lies (O'Sullivan et al., 2009). In particular, if O'Sullivan's reasoning is correct, then our predictions will fail when we try to replicate our findings with older, more experienced adults judging our high-stakes truths and lies.

As an initial nonstudent sample, we chose university professors for Experiment 3. Because our lies involved cheating college students, we reasoned that university professors would have everyday professional experience with student lies about cheating. Some of the studies claiming to identify experts or detection wizards are conducted with samples from complementary sender and judge populations (i.e., accused criminals and law enforcement professionals). Because the senders in our experiments are all students, using a sample of professors provides a reasonable analog of the criminal-sender, law enforcement-judge condition.

It was further reasoned that additional strong and convincing evidence for the senders-matter-but-judges-are-interchangeable proposition might be obtained from an intercultural replication. To test if nonnative English-speaking judges from outside North America might see the tapes in the same way as the North American judges in Experiments 1–3, university students from a university in Seoul, South Korea were recruited for Experiment 4.

In Experiment 5, the judges were highly trained interviewers, investigators, and polygraphists from a U.S. government security and intelligence agency. Although these results are used with permission, the specific agency cannot be identified. Each of these judges had substantial training in lie detection. A subset of the sample consisted of long-time professionals with 15 or more years of experience.

Experiment 3: Method

The participants in Experiment 3 were 30 faculty members at a different Midwestern U.S. university than that of the students in Experiments 1 and 2. The sample was predominantly male (76.7%), on average 55.4 years old ($SD = 9.5$), and had, on average, 19.3 years of teaching experience ($SD = 13.4$). Sixteen (53.3%) reported having previously caught student cheaters. However, neither years of teaching experience nor cheater-catching experience correlated with truth bias or detection accuracy.

Unlike the other experiments, the data in Experiment 3 were collected online. Participants were provided with a Web link and completed the research on computers of their choice. After instructions and informed consent, they saw and judged, in random order, each of the 20 unique videotaped interviews used in the previous experiments. Each tape was presented one at a time, and after viewing each tape, participants were asked to make a truthful noncheater or a lying cheater determination. The resulting judgments were then scored into the experimental conditions used in the previous experiments. The format of Experiment 3 also allowed scoring accuracy and truth bias into a 2 (actual truth–lie) by 2 (sincere–insincere demeanor) presentation format.

Results of Experiment 3

In the professor data, participants correctly identified 59.7% of the interviews. This value was statistically larger than both the 50% chance rate, $t(29) = 5.40$, $p < .001$, and the 54% meta-analysis finding, $t(29) = 3.17$, $p = .004$. Participants also tended strongly toward truth bias, judging 66.0% of messages as honest; the proportion of messages judged as honest was statistically greater than 50%, $t(29) = 4.34$, $p < .001$ (two-tailed).

In the demeanor–veracity matched condition, accuracy (78.2%, $SD = 18.2$) was well above both chance, $t(29) = 8.49$, $p < .001$, and meta-analysis levels, $t(29) = 7.28$, $p < .001$. In contrast, in the demeanor mismatched condition, accuracy (40.7%, $SD = 13.9$) was well below chance, $t(29) = -3.65$, $p = .001$. Accuracy in the two conditions was statistically and substantially different, paired samples $t(29) = 8.00$, $p < .001$. Truth bias in the two conditions was also statistically different, $t(29) = -4.63$, $p < .001$.

In the sincere demeanor condition, truth bias (85.0% honest, $SD = 24.4$) was substantial and well above 50%, $t(29) = 7.85$, $p = .001$. In contrast, lie bias (47.0% honest, $SD = 25.4$) was observed in the insincere demeanor condition. However, the proportion was not significant below 50%, $t(29) = -0.64$, $p = .53$. The two conditions differed substantially in the proportion of messages judged as honest, $t(29) = 7.87$, $p < .001$. The two conditions were also different in terms of accuracy, $t(29) = -4.78$, $p < .001$.

Findings of the faculty data are summarized in Table 3. It is notable that, for this group of judges, demeanor made a larger difference when judging liars than when judging truthtellers. This trend was also evident in each of the previous and subsequent experiments.

Experiment 4: Method

The participants in Experiment 4 were $N = 182$ (40% men, mean age = 22.3) students from a university in Seoul, South Korea. Experiment 4 used an independent groups design identical to Experiment 2 except that the tapes included two practice interviews, the responses to which were not included in the analysis. The survey, including instructions, truth–lie items, questions regarding ability to understand English, and demographic items, was translated into Korean, but the tapes of senders were shown in English without subtitles or other translation.

As a validity check, participants were asked to rate their ability to understand English on a 4-item scale ($\alpha = .89$). Self-reported language ability did not differ across conditions, $F(3, 178) = 1.91$, $p = .13$, $\eta^2 = .03$, and was not correlated with either accuracy, $r = -.06$, $p =$ ns, or truth bias, $r = .06$, $p =$ ns.

Results of Experiment 4

In the Korean data, participants correctly identified 55.2% of the interviews. This value was statistically larger than the 50% chance rate, $t(181) = 3.32$, $p = .001$. Participants were also significantly truth biased, judging 60.8% of messages as honest, $t(181) = 8.15$, $p < .001$.

Table 3 Accuracy and Truth Bias in Experiment 3 (University Faculty; $N = 30$)

	Accuracy		Percent Honest	
	M (%)	*SD* (%)	*M* (%)	*SD* (%)
Tape				
Demeanor–veracity matched	78.2	18.2	58.1	19.4
Demeanor–veracity mismatched	40.7	13.9	74.1	24.9
Sincere demeanor	51.6	10.2	85.0	24.4
Insincere demeanor	67.9	16.0	47.0	25.4
Across tapes	59.7	9.8	66.0	20.2
Veracity–demeanor				
Honest–sincere demeanor	86.3	21.8	86.3	21.8
Liar–insincere demeanor	70.5	30.0	29.5	30.0
Honest–insincere demeanor	64.8	29.7	64.8	29.7
Liar–sincere demeanor	16.7	27.3	83.3	27.3
Across tapes	59.7	9.8	66.0	20.2

Note: University faculty viewed each interview once using an online questionnaire; interview order was fully randomized. Judgments were grouped according to base veracity–demeanor and matched to the videotape conditions in the other experiments. This was not the case in the other experiments where each interview was judged twice.

In the demeanor–veracity matched condition, accuracy (70.7%) was well above chance, $t(42) = 7.21$, $p < .001$. In contrast, in the demeanor–veracity mismatched condition, accuracy (33.9%) was well below chance, $t(44) = -8.84$, $p < .001$. Accuracy in the two conditions was statistically and substantially different, $F(1, 85) = 117.62$, $p < .001$, $\eta^2 = .59$. Truth bias in the two conditions was also statistically different, $F(1, 85) = 14.80$, $p < .001$, $\eta^2 = .15$.

In the sincere demeanor condition, truth bias (64.3% honest) was substantial and well above 50%, $t(46) = 5.52$, $p < .001$. Truth bias (51.3% honest) did not differ from chance in the insincere demeanor condition, $t(47) = 0.55$, $p =$ ns. The two conditions differed substantially in the proportion of messages judged as honest, $F(1, 93) = 14.43$, $p < .001$, $\eta^2 = .13$. The two conditions also differed in terms of accuracy, $F(1, 93) = 7.15$, $p < .01$, $\eta^2 = .07$. The findings of Experiment 4 are summarized in Table 4.

Experiment 5: Method

The fifth experiment involved $N = 35$ professionally trained and currently practicing deception detection experts employed by a U.S. security and intelligence agency. The sample of experts was 57.6% female, and the age distribution was positively skewed with $M = 31.6$, $SD = 9.4$. The amount of professional experience was also positively skewed and ranged from 6 months to more than 30 years with a mean of 7.7 years ($SD = 8.4$). The distribution of professional experience was such that most of the sample (80%) reported 7 or less years of experience, but seven individuals (20%)

Table 4 Accuracy and Truth Bias in Experiment 4 (Korean Students; $N = 182$)

	Accuracy		Percent Honest	
Tape	*M* (%)	*SD* (%)	*M* (%)	*SD* (%)
Demeanor–veracity matched	70.7	16.8	57.4	16.8
Demeanor–veracity mismatched	33.9	12.2	70.7	15.3
Sincere demeanor	62.5	19.0	64.3	17.7
Insincere demeanor	55.2	13.8	51.3	15.7
Across tapes	55.2	21.0	60.8	17.8

reported 15 or more years of experience. Because no individual reported between 8 and 14 years of experience, the 7-or-less years and 15-or-more years distinction provided a natural cut point for distinguishing more and less experienced experts in this sample. The data were collected as part of a professional training seminar. The training was provided at no cost in exchange for the experts' participation in research.

Each of the experts judged a series of 32 interview tapes that were selected from the 44 tapes used in the Levine et al. (2010) direct questioning condition. Imbedded within the 32 tapes judged were 14 of the 20 interviews used in the previous four experiments. The 14 interviews contained seven veracity–demeanor matched interviews (three truths and four lies) and seven veracity–demeanor mismatched interviews (two truths and five lies). The experts' truth and lie judgments on the relevant 14 interviews were scored for accuracy as a conceptual replication of the previous findings with a sample of experts.

Results of Experiment 5

Overall, the accuracy for the experts was 65.3%. This value was statistically greater than the 50–50 chance rate, $t(43) = 11.65$, $p < .0001$, and the 54% meta-analysis mean, $t(43) = 8.60$, $p < .0001$.

Accuracy for the demeanor–veracity matched senders was 96.3%, a level of accuracy far above chance, $t(34) = 37.96$, $p < .0001$. In contrast, accuracy for the demeanor–veracity mismatched senders was 34.3%, a level much below chance, $t(34) = -5.97$, $p < .001$. The difference between the two sets of senders was statistically significant and large, $F(1, 34) = 387.73$, $p < .0001$, $\eta^2 = .87$.

Among demeanor–veracity matched senders, accuracy was high for both sincere-looking honest senders ($M = 94.3\%$) and insincere-looking liars ($M = 97.9\%$). For mismatched senders, accuracy was near chance ($M = 51.4\%$) for the relatively insincere-appearing honest senders, but much below chance ($M = 27.4\%$) for sincere-acting liars, $t(34) = -8.65$, $p < .0001$.

Years of experience were negatively correlated with total accuracy ($r = -.43$, $p < .02$), accuracy for mismatched senders ($r = -.54$, $p < .001$), and accuracy for sincere liars ($r = -.43$, $p < .02$). Experience was positively associated, although not statistically significantly so, with accuracy on matched senders ($r = +.25$, $p = .16$).

Because 7-or-less years experience and 15-or-more years experience provided a natural gap in the distribution of experience, the experts were split into less experienced and more experienced groups at the point of this gap in the distribution. The less–more years of experience split significantly moderated the main effect for demeanor matching on accuracy, $F(1, 30) = 8.38$, $p < .01$, $\eta^2 = .02$, although this interaction effect was small compared with the much larger demeanor main effect. The more experienced experts with 15 or more years of experience obtained perfect 100.0% accuracy on the demeanor–veracity matched senders including both sincere-acting honest senders and the insincere-looking liars. However, they obtained only 20.4% accuracy on the mismatched senders (insincere honest = 35.7% and sincere liars = 14.3%). In fact, the matched–mismatched induction explained 98% of the variance in accuracy with the more experienced experts, $F(1, 6) = 760.50$, $p < .001$, $\eta^2 = .98$. The matched–mismatched effect was less dramatic with the less experienced experts, $F(1, 24) = 238.72$, $p < .0001$, $\eta^2 = .83$. For the less experienced experts, veracity matched accuracy was $M = 95.4\%$ and mismatched accuracy was $M = 37.4\%$. Thus, expertise increases strength of sender demeanor effects. Demeanor has a greater effect on experts than nonexperts and stronger effects on experts with more experience than experts with less experience. Accuracy results of Experiment 5 are summarized in Table 5.

Discussion of Experiments 1–5

The results of the first five experiments demonstrated that manipulating sender demeanor creates predictable and large differences in detection accuracy. Subjects who saw the demeanor–veracity matched senders, which included both the sincere-acting truthtellers and the insincere-acting liars, were accurate at rates between 70.7% (Experiment 2) and 100% (most experienced experts, Experiment 5). In each case, accuracy was well above meta-analysis levels (54%). In contrast, participants rating the demeanor–veracity mismatched senders, which included both insincere

Table 5 Accuracy in Experiment 5 (U.S. Government Agents; $N = 35$)

	Accuracy			By Experience (M)	
Condition	M (%)	SD (%)	95% CI	≤ 7 years (%)	>15 years (%)
Demeanor–veracity matched	96.3	7.2	±2.3	95.4	100.0
Demeanor–veracity mismatched	34.3	15.6	±5.1	37.4	20.4
Honest–sincere demeanor	94.3	15.1	±5.1	93.3	100.0
Liar–insincere demeanor	97.9	7.1	±2.4	97.0	100.0
Honest–insincere demeanor	51.4	30.9	±10.2	54.0	35.7
Liar–sincere demeanor	27.4	18.2	±6.1	30.4	14.3

Note: In the honest conditions, the percent judged honest is the same as the percent accurate. In the lie conditions, the percent judged honest is 1.0 − accuracy.

truthtellers and sincere liars, were correct at rates significantly below chance (50%) level (20.4–41.4%). In terms of raw accuracy, the demeanor induction created between a 36.3% and a 79.6% swing in raw accuracy and across the first five experiments accounted for between 59% and 98% of the variance in accuracy. The consistency of the matched–mismatched induction of detection accuracy in the five experiments is graphed in Figure 1.

While the sender demeanor induction produced large and consistent differences, the findings also demonstrate relative judge constancy. Comparatively, there was little within-cell variance across studies. Nonstudent and Korean judges yielded the same results as the U.S. college students in the first two experiments. Furthermore, as individual differences in judge ability would contribute to the error term beyond sampling and measurement error and as the sender induction explained more than 50% of the variance in each of the experiments, the variance attributable to all other causes combined, including judge ability, was less than that of the sender demeanor induction.

Demeanor effects were the strongest in Experiment 5 with U.S. government agency experts. Among these experts, the demeanor effects were more pronounced with the more experienced subset of experts. This suggests that experts are more sensitive to demeanor than judges who have not been trained in deception detection. When demeanor was consistent with actual honesty, the experts achieved accuracy

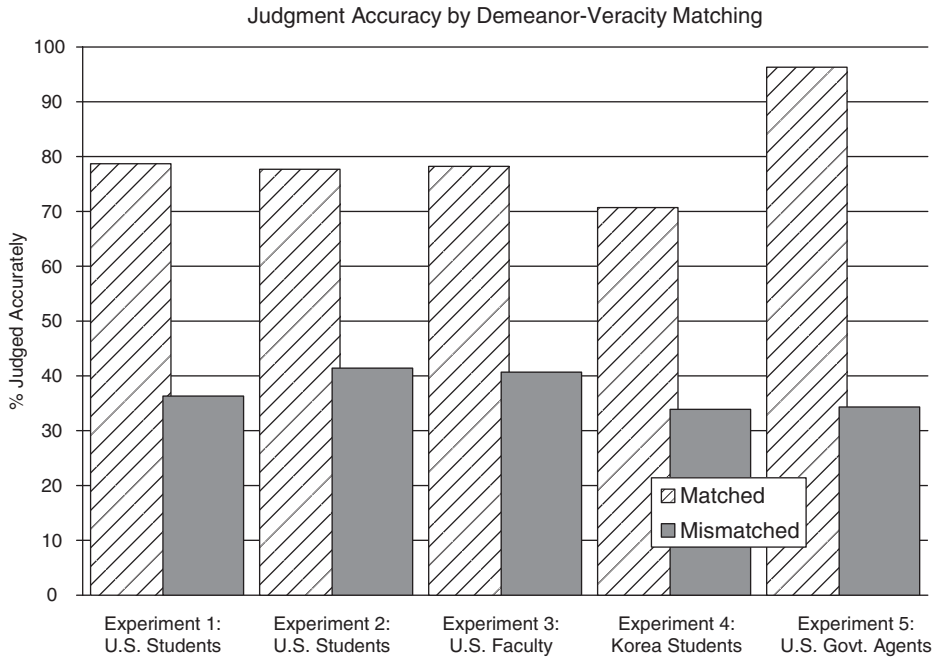

Figure 1 Consistency of accuracy findings in Experiments 1–5.

rates in excess of 90%. Thus, when demeanor was matched with veracity, it helped experts more than the nonexpert judges in the previous experiments who had levels of accuracy below 90% in the demeanor matched condition. However, especially for liars, when demeanor was mismatched, experts were misled by their reliance on demeanor. This was particularly true for the smooth liars who fooled the most experienced experts more than 85% of the time.

The complete independence of the demeanor induction and actual honesty prevailed in Experiments 1 and 2 but was not fully maintained in Experiments 3 and 4 where the matched–mismatched conditions differed in truth bias and the sincere–insincere conditions differed in accuracy. The intended effects of the demeanor induction were more consistent and invariably larger in effect size than the effects implying demeanor–veracity nonindependence. Nevertheless, the data were not consistent with uniform empirical independence.

In Experiments 3 (university professors) and 5 (U.S. government agents), veracity was a strong moderator of demeanor effects. Demeanor made a larger impact on accuracy when the sender was lying. In Experiment 3, while the demeanor induction made a 20% difference in raw accuracy for honest senders, it made a whopping 54% difference for liars. Liars presenting a dishonest demeanor were correctly identified 71% of the time. In contrast, the liars with an honest demeanor were correctly identified only 17% of the time. This pattern was also observed in Experiment 5. One possible explanation for the smaller demeanor effects for honest senders is the nature of the questioning. Levine (2007) designed the questions to improve accuracy, and the ''what will your partner say'' final question may have helped honest senders, despite their demeanor, to convey their honesty (Levine et al., 2010). Inconsistency is often seen as evidence of dishonesty; thus the use of three questions provided the judges with multiple opportunities to catch dishonest senders.

Demeanor cue identification studies

Study 6

The previous five experiments document that individual differences across senders have a large impact on deception detection accuracy. The findings from the five experiments, however, beg the question of what senders are doing that make them more or less believable.

The literature has identified several behaviors that are linked with truth and lie judgments. For example, Zuckerman et al. (1981) suggest that gaze, smiling, fewer posture shifts, short response latencies, and speech that is more fluent would be associated with a sincere demeanor. Bond and the Global Deception Research Team (2006) found that acting nervous, providing incoherent answers, and, especially, gaze aversion were believed to be associated with deception worldwide. Study 6 and the two subsequent studies investigate whether these and other behaviors contribute to the consistent sincere and insincere demeanor effects reported in the five experiments.

Participants, procedures, and results

The participants were the same 30 students as in Experiment 1, and Study 6 took place during the next class session. The participants were provided with the results of Experiment 1, and they were told it would be useful to know on what they had based their judgments. The tapes were played several times and the participants were encouraged to discuss what made each sender seem believable or suspicious. The instructor–researcher took notes, and the tape viewing and discussion continued until all ideas were exhausted and the participants were confident that they had identified all important demeanor cues.

The ideas that emerged over the viewing of several senders were culled and grouped by the first author to form the list of 11 demeanor impressions and behaviors presented in Table 6. Some of the items are more gestalt impressions such as appearing composed, friendly, or nervous. Others are more specific, such as gaze aversion or fidgeting. The list encompasses body movements, tone and voice, and verbal content. Generally, the list is consistent with expectations based on the previous literature (e.g., Bond & The Global Deception Research Team, 2006; Zuckerman et al., 1981).

Study 7

The next study was designed to test whether the demeanor cues identified in Study 6 varied between senders known to differ in demeanor and whether the demeanor cues were associated with honesty judgments.

Table 6 Eleven Behaviors and Impressions Linked to Honest–Dishonest Demeanor

Sincere (Honest) Demeanor Cues
1. Confidence and composure
2. Pleasant and friendly interaction style
3. Engaged and involved interaction style
4. Gives plausible explanations
Insincere (Dishonest) Demeanor Cues
5. Avoids eye contact
6. Appears hesitant and slow in proving answers
7. Vocal uncertainty (conveys uncertainty in tone of voice)
8. Excessive fidgeting with hands or foot movements
9. Appears tense, nervous, and anxious
10. Portrays an inconsistent demeanor over course of interaction
11. Verbal uncertainty (conveys uncertainty with words)

Note: The honest demeanor index for an individual sender consists of subtracting the average 1–7 rating of the last seven items from the average 1–7 rating of the first four items. These 11 behaviors, as an index of honest–dishonest demeanor, were created and copyrighted by the first author and are considered proprietary. Use of the materials in Table 6 in scales, behavioral coding, training, and other applications is permissible only with the prior written permission of the first author, who owns the intellectual property rights and copyright to this demeanor index. This table is provided here with the permission of the author.

Study 7: Participants and procedures

The participants were the same $N = 113$ students who participated in Experiment 2 and were drawn from a subject pool at the same large Midwestern U.S. university as the previous study. Study 7 was conducted subsequent to Experiment 2. Participants again watched the video tapes of each of the 20 senders used in the previous experiments. Instead of being judged on honesty, however, each sender was rated on each of the 11 demeanor cues generated in Study 6. Each cue was rated on a 7-point scale for how well it described each sender's communication style; a higher score indicated better fit between a cue and the communication style. Mean honesty judgments for each sender were taken from the demeanor matched and demeanor mismatched conditions of Experiment 2 and correlated with the behavior ratings.

Results and discussion of Study 7

The demeanor cues were scored three ways. First, each cue was scored individually. Second, they were scored as averages on sincere ($\alpha = .83$) and insincere ($\alpha = .89$) subdimensions. Finally, they were scored as a unidimensional index ($\alpha = .87$) with the insincere score subtracted from the sincere score.[2] An exploratory factor analysis with principal axis extraction and promax rotation yielded three eigenvalues greater than 1.0. In the three-factor solution, gaze loaded on its own factor. Examination of the scree plot suggested using one- or two-factor solutions. In the two-factor solution, the sincere items loaded on one factor, the insincere items loaded on the other, and the correlation between two factors was $r = -.59$ ($-.69$ corrected for attenuation). Examination of the correlations in Table 7 suggests that the ratings of all individual cues and the two subdimensions are parallel with respect to the demeanor induction, honesty judgments, and actual sender veracity. Therefore, it is concluded that the index can meaningfully be scored as a unidimensional measure of honest–dishonest demeanor.

Because each judge rated multiple senders, it was possible to examine both judge and sender effects. The judge effect on perceived demeanor was $F(112, 1016) = 2.03$, $p < .001$, intraclass correlation (ICC) $= .09$, $\eta^2 = .18$ and the sender effect was $F(19, 1109) = 74.46$, $p < .001$, ICC $= .56$, $\eta^2 = .56$. Multilevel modeling analysis yielded similar conclusions but failed to provide additional insights and is therefore not reported.

The behavior ratings differed substantially for the senders comprising the sincere and insincere demeanor inductions in the first five experiments: sincere subscale, $F(1, 1127) = 182.23$, $p < .001$, $\eta^2 = .14$; insincere subscale, $F(1, 1128) = 2{,}118.60$, $p < .001$, $\eta^2 = .26$; and total demeanor index, $F(1, 1127) = 6{,}091.80$, $p < .001$, $\eta^2 = .25$. Correlations between the honesty ratings from Experiment 2 and the demeanor index were sincere $r = .55$, insincere $r = -.52$, and total $r = .60$. The correlation between the percentage of judges rating a sender as honest and the same sender's total demeanor score was $r = .52$ for honest senders and $r = .63$ for liars. Correlations are provided in Table 7.

Table 7 Correlations Between Demeanor Cues and Honesty Judgments (HJ), the Demeanor Induction (DI), and Actual Sender Veracity (AV) in Studies 6 and 7

	Study 6 ($N = 113$ Judges)			Study 7 ($N = 93$ Senders)	
	HJ	DI	AV	HJ	AV
Plausible explanation	0.43*	0.14*	0.45*	na	na
Maintained eye contact	0.31*	0.38*	0.16*	0.27*	−0.08
Confident and composed	0.52*	0.46*	0.38*	0.35*	0.02
Pleasant and friendly	0.45*	0.29*	0.49*	0.34*	0.18
Engaged and involved	0.43*	0.25*	0.47*	0.28*	0.06
Hesitant	−0.49*	−0.48*	−0.31*	−0.27*	0.04
Vocal uncertainty	−0.51*	−0.46*	−0.27*	−0.34*	−0.05
Fidgeting	−0.19*	−0.24*	−0.10*	−0.12	−0.11
Nervous	−0.42*	−0.28*	−0.33*	−0.25*	0.01
Demeanor change	−0.34*	−0.36*	−0.16*	−0.41*	0.20*
Verbal uncertainty	−0.48*	−0.48*	−0.29*	−0.39*	−0.14
Sincere subtotal	0.55*	0.37*	0.51*	0.41*	0.06
Insincere subtotal	−0.52*	−0.51*	−0.31*	−0.38*	−0.10
Demeanor index total	0.60*	0.50*	0.45*	0.44*	−0.09

(*$p < .05$.)

Study 8

The results of the previous study (a) suggest that the 11 cues identified in Study 6 are a unidimensional index, (b) show that behavioral ratings on that index were predicted from differences between the senders who formed the sincere and insincere conditions in the previous experiments, and (c) also show that the behavioral ratings were also correlated highly with honesty judgments from Experiment 2. Thus, the scores function as a unidimensional index of honest–dishonest demeanor and seem to indicate the set of cues judges use when making deception judgments. Study 8 cross-validated these findings with different judges and senders, and independent scoring by trained coders.

Study 8: Participants and procedures

The new senders were participants in a cheating experiment similar to Levine (2007). In all, 104 students participated; 11 individuals who cheated also confessed under questioning and were included as honest senders (by virtue of their confessions). Differences between the new sender tapes and the Levine (2007) tapes included a different interviewer and four different question sets. The question set variable, however, did not significantly affect any of the variables under consideration here.

Each of the 104 tapes was coded independently for 10 of the 11 demeanor cues by each of seven trained and paid undergraduate coders. The plausibility of explanation cue was not coded because some question sets did not require explanations. Similar

to Study 6, coders used 1–7 scales to rate each demeanor cue. Intercoder reliability exceeded $\alpha = .70$ for all 10 cues. Scores for each behavior were then averaged across the coders, and the cues were scored individually, for the sincere and insincere subscales, and as a unidimensional index.

Honesty ratings were made by a separate sample of $N = 157$ undergraduate judges recruited from a communication department research pool. As a result of the number of senders and interview length, each judge evaluated only 26 of the senders. Each sender was evaluated by between 35 and 46 judges; percent judging the sender honest and percent accuracy averaged across judges were recorded for each sender.

Results and discussion of Study 8

The correlations between the demeanor coding, judges' honesty ratings for the same sender, and with the sender's actual veracity were examined. The results are provided in Table 7. Nine of the 10 demeanor cues were significantly correlated in the predicted direction with judge honesty ratings (fidgeting was not significantly correlated with honesty). Only one of the 10 demeanor cues was correlated with actual veracity (demeanor change over time was negatively correlated; i.e., liars are more likely to manifest a demeanor change). The demeanor index total was not significantly associated with actual honesty, $r = -.09$, but the demeanor index total was correlated with honest judgments at $r = .44$. The correlation was lower, but still statistically significant, for honest senders only, $r(91) = .41$, $p = .002$. For actual liars only, the correlation was $r(9) = .86$, $p = .001$.

Plots of the demeanor–honest judgment relationships for truthtellers and liars are provided in Figure 2. The plots demonstrate the substantial variance in sender demeanor, the association between demeanor ratings and honesty judgments, the moderating effects of actual veracity, and the independence between demeanor ratings and accuracy.

General discussion

Traditional (Ekman & Friesen, 1969) and contemporary (O'Sullivan, 2008) leakage theory, Zuckerman et al.'s (1981) four-factor theory, and Buller and Burgoon's (1996) interpersonal deception theory all specify that (a) there are different behaviors associated with truth and deception and that (b) skilled message recipients should therefore be able to reliably distinguish truths from lies based on observable sender behavior (demeanor). The repeated finding that most people are not very good at detecting lies most of the time appears inconsistent with these theories. Although auxiliary propositions can be invoked to save the theories from the data by asserting a network of methodological limitations explaining why the vast majority of studies fail to provide support (e.g., Frank & Feeley, 2003; O'Sullivan, 2008), a more parsimonious explanation might be that these theories are mostly wrong. Sender demeanor may not provide a useful way to distinguish truth from lie for most senders (Levine, 2010). This series of studies investigated whether or not some new insights

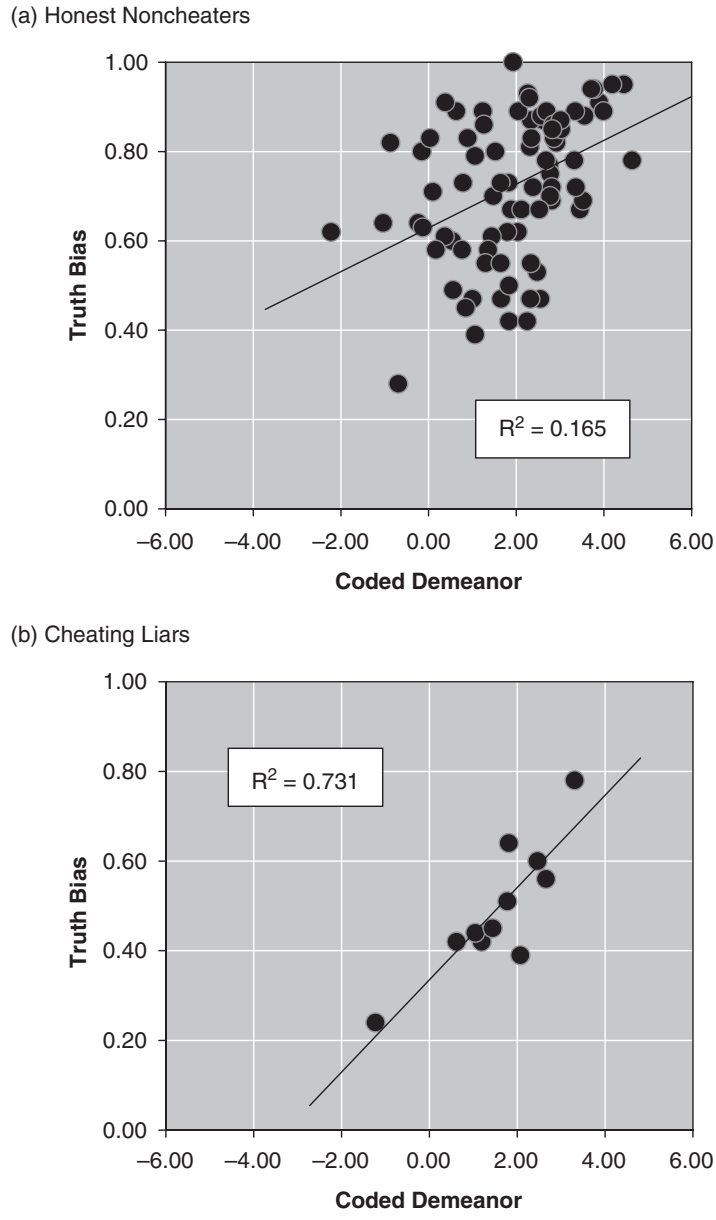

Figure 2 Plots of coded demeanor behaviors onto the proportion believed by untrained judges for honest noncheaters ($n = 82$) and lying cheaters ($n = 11$) in Study 8. (a) Honest noncheaters; (b) cheating liars.

Note: Truth bias is the proportion of judges who believe a given sender. For honest senders, truth bias = accuracy, whereas for liars, truth bias = 1 − accuracy. Positive values for coded demeanor indicate that the sender exhibited sincere behaviors.

might be gained by presuming that sender demeanor was (mostly) independent of deceptive intent.

We do not doubt that leakage and deception cues can and do exist, and we do not believe that demeanor and honesty are completely uncorrelated (i.e., $r = .000$). Accuracy in deception detection experiments is statistically better than chance. Absent any behavioral signal, this would not be the case, and the meta-analysis mean would coincide with chance-level accuracy. It does not. Levine (2010) proposed the concept of a few transparent liars to explain the statistically significant but only slightly better than chance accuracy findings that typify the literature. The idea was that if leakage and deception cues applied to some small percentage (e.g., 10%) of senders, then the accuracy findings in the literature make sense. The few leaky liars would have obvious consistency between demeanor and veracity and consequently most judges would get them right thereby doing much better than chance. But, because there are only a few leaky liars, for the majority of senders behavioral displays are not meaningfully linked with honesty. The result is the low-accuracy ceiling typical of most of the literature.

The results presented here showing that sender demeanor is mostly independent of actual veracity is consistent with Levine (2010). The implications are seen in the following series of propositions:

1. If accuracy and demeanor are substantially correlated for only a few leaky liars, then demeanor and honesty will be unrelated for a majority of senders.
2. Senders vary considerably in demeanor; this follows directly from Bond and DePaulo's (2008) recent meta-analysis.
3. Furthermore, it is the sender, not the judges, driving the accuracy of the deception judgment process. This also follows from Bond and DePaulo (2008) and is validated in Levine et al. (2010) and Experiments 1–5 and 8 of this research.

Putting these three propositions together leads to the conclusion that:

4. The variation in sender demeanor is mostly independent of actual honesty.
5. And as a consequence, there is a ceiling on how much accuracy can be obtained when honesty judgments are based on sender demeanor; this latter conclusion is consistent with the findings of the Bond and DePaulo (2006) meta-analysis.

Putting these propositions and conclusions together further leads to the predictions that:

6. Judge accuracy can be manipulated by selectively controlling which senders are rated as demonstrated by the first five experiments in this report, and
7. That varying which senders are judged makes much more difference than selecting which judges do the judging. This prediction is consistent with the lack of judge effects observed in meta-analyses (Aamodt & Custer, 2006; Bond & DePaulo, 2006, 2008) and was shown in Experiments 1–5 of this research.

These predictions were tested by experimentally varying senders and by quasi-experimentally varying judges across five experiments. Each sender was one of the four types: honest-appearing (sincere demeanor) truthtellers, deceptive-appearing (insincere demeanor) truthtellers, honest-appearing liars, or deceptive-appearing liars. The judges were college students at a large Midwestern U.S. university, professors from a different U.S. university, Korean college students at a university in South Korea, or U.S. government security and intelligence agents. As anticipated, varying senders made a huge difference, explaining between 59% and 98% of the variance in detection accuracy. Using different judges—whether the judges were American students, much older adults (ostensibly experienced in judging student honesty), people from different cultures and countries, or security agents trained to identify criminal and subversive behavior—made little difference as shown in Figure 1. The effect of the sender induction on accuracy proved robust across judges. When veracity and demeanor were matched, accuracy was well above both chance and meta-analysis levels in all five experiments. When veracity and demeanor were mismatched, accuracy was consistently, significantly, and substantially below chance; and, in fact, the lowest accuracy in detecting sincere liars was observed for the most experienced and trained judges. If demeanor was not largely independent of veracity, this induction and these findings would not be possible. But, presuming that demeanor and veracity were independent encouraged us to try to manipulate them independently and to ultimately explain as much as 98% of the variance in accuracy scores.

While the sender demeanor experiments provide a compelling demonstration of the powerful impact of demeanor, these findings alone would have been unsatisfying because they provide little insight into what behavioral displays contribute to an honest or dishonest demeanor. Therefore, the final three studies investigated what makes some people appear sincere and others duplicitous, independent of actual honesty. First, we had the participant judges from the first detection experiment watch the senders again (and again, repeatedly) and try to articulate what was guiding their judgments. On the basis of cues identified in those data, we had a second set of subjects rate the senders for demeanor-linked behaviors and the associations between behavior ratings and honesty judgments reported in a previous experiment were examined. Finally, we cross-validated the behavioral cues identified using different judges and senders. This produced a set of 10 or 11 cues—ranging from overall composure and interaction style to specific acts such as eye contact avoidance, fidgeting, and nervous tension—that form a unidimensional index of honest–dishonest demeanor, that (a) are highly and consistently related to honesty judgments and (b) are not highly correlated with actual honesty. These findings depart from those of Frank and Ekman (2004), who found that believability was conveyed primarily through dynamic facial behavior.

Examination of Figure 2 shows (a) considerable variance in the percentage at which individual senders are believed, (b) considerable variance in the sender's demeanor scores, (c) the variance in sender believability and coded demeanor occurs

for both truthful and deceptive senders, and (d) that coded demeanor and the percentage of judges who believe a sender covary predictably. Honest senders were believed by anywhere between 28% and 100% of judges, whereas liars were believed by between 24% and 78% of the judges. Honest senders' demeanor scores ranged from −2.23 to 4.64, whereas liars' scores ranged from −1.23 to 3.31. These reflect an almost 7-point swing on a 13-point index. The correlations between the percentage of judges who believed a given sender and the coded demeanor of the same sender were $r = .41$ for honest senders and $r = .86$ for liars. The second correlation is especially impressive because the correlation is larger than the measurement reliability, suggesting a near-perfect association when adjusting for measurement error. In short, regardless of honesty, senders vary considerably both in behavioral cues linked with demeanor and in the extent to which they are believed by judges, and liars who engage in behaviors linked with an honest demeanor and avoid the behaviors linked with a dishonest demeanor tend to be believed substantially more often than those who do otherwise.

These findings provide a parsimonious and compelling explanation for the accuracy ceiling observed in the literature. Because senders differ considerably in how honest they appear and because an honest appearance rests more on individual differences that are independent of actual honesty than on deception-linked behaviors, judges who rely on demeanor will get a substantial proportion of judgments wrong, and this is especially true when they are judging liars. That is, because a number of honest senders present behaviors that tend to be judged as dishonest and because a number of liars present themselves as honest acting, errors in judgment are assured. Furthermore, judges (regardless of age, culture, training, or enrollment in Midwestern U.S. universities) tend to look for the same set of demeanor cues and consequently evaluate senders in the same way as other judges. This means that the same errors tend to be made by different judges, and this accounts for the larger variance in senders than for judges that was observed in meta-analysis (Bond & DePaulo, 2008). It also accounts for the persistence of the accuracy ceiling across decades of deception detection research involving numerous research design modifications, all of which made little difference.

Individual differences in sender honest demeanor may be related to individual differences in the proclivity to lie (Levine, 2010; Serota, Levine, & Boster, 2010). Because lying is volitional and because there are often sanctions associated with being caught in a lie, people with more honest demeanors may be more likely to lie because their lies are more likely to be successful and avoid detection. Preliminary findings consistent with this thinking were recently reported by Levine et al. (2010). The implication is that variance in sender demeanor coupled with random assignment tends to push accuracy rates toward chance, but variance in sender demeanor coupled with self-selected deception provides a mechanism for below-chance accuracy.

Perhaps, one of the most interesting and important findings here is that the 10 or 11 behaviors linked with sender honest demeanor appear to form a unidimensional index, and this appears true both in the untrained rater data of Study 7 and in the trained coder data from Study 8. The demeanor cues were all substantially

intercorrelated; they were associated in a parallel manner with the demeanor induction and judge honesty ratings. Few nonredundant conclusions were gained by scoring them separately or as sincere and insincere subdimensions. The implication of unidimensionality is that these behaviors tend to be seen by observers as a package or a whole. Although virtually all of the cues have been linked with perceived honesty in the previous research, what is new is the idea of adding them up as a unidimensional index for honest demeanor that is linked with believability rather than actual honesty.

A second finding is that demeanor effects appear to be moderated by veracity such that demeanor effects are larger for liars than truthtellers. The effect was evident in Experiments 3 and 5 and in Study 8. In Experiments 3 and 5, the demeanor induction created a larger difference in accuracy between insincere and sincere liars than insincere and sincere truthtellers. In Study 8, the same effect was reflected in the stronger association between coded demeanor and honest judgments for liars than truthtellers.

Levine et al. (2010, see Table 2), reported small differences between the most believable liars and the most believable truthtellers but much greater differences between the least frequently believed liars and truthtellers. Truthtellers were believed between 34.4% and 100% of the time, and only three honest senders were believed by fewer than 65% of judges. Liars were believed between 7.8% and 95.3% with 8 of 22 below 40% believability and 9 of 22 above 70% believability. These results show substantially more variance in believability for liars than truthtellers (see also Frank & Ekman, 2004, Figure 1, for similar findings). Simply put, a few liars are very leaky (cf. Levine, 2010), whereas no truthtellers were as poorly demeaned as these exceptionally leaky liars. The few leaky liars created more variance in believability for liars than truthtellers, and the greater variance in believability for liars allows demeanor to have proportionally larger impact. This study appears consistent with this critical distinction in deception research.

Perhaps the most important implication of these findings is that if naturally occurring variation in sender demeanor produces the accuracy ceiling in deception detection, then breaking through the ceiling to obtain higher accuracy will likely involve basing judgments on factors other than sender demeanor. Along these lines, high levels of accuracy have recently been reported with the strategic use of evidence (Hartwig et al., 2006) and content in context (Blair et al., 2010) approaches. It is noteworthy that neither of these approaches focuses on nonverbal or linguistic cues to deception and neither is guided by theoretical models involving nonstrategic nonverbal leakage, deception-based arousal, or cognitive effort. Instead, both approaches ignore sender demeanor in favor of evaluating message content in light of prior knowledge based on evidence or context. This series of studies suggests that if the goal is to increase accuracy, then abandoning the ideas of detecting deception through sender demeanor may be fruitful. If, however, the goal is to understand why some people are believed more often than others, then sender demeanor, and the behavioral amalgamation of cues that comprise it, appears to be the primary factor in operation.

Acknowledgment

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

Notes

1 Frank and Ekman (2004) examined truth judgments made of a small number of senders ($N = 15$) across two different truth-deception tasks. Individual senders who were more frequently believed in one task were more often believed by different judges in the second task ($r = .87$). High correlations were obtained for both truths and lies ($r = .89$ and .85, respectively). Consistent with this argument and Zuckerman et al.'s (1979) demeanor bias, Frank and Ekman observed large differences in believability from sender to sender (see Frank & Ekman, Figure 1, p. 489). While Frank and Ekman focused on showing that sender demeanor was consistent across two situations, this article emphasizes generality across judges and explores the implications of these large and consistent sender effects for deception detection accuracy. In addition, Frank and Ekman attributed sender differences primarily to dynamic facial behaviors while this research points to a broader array of behaviors that function as a package to form a perceptual gestalt.

2 When scoring the index as a unidimensional measure of sincere–insincere demeanor, scores on the two subdimensions were first averaged, then the insincere average score was subtracted from the sincere average score, creating a 13-point index ranging from −6 to +6. An advantage of this scoring is that the resulting values form an intuitive metric with zero as the midpoint, with negative scores reflecting degrees of insincerity and positive values reflecting increasing sincerity. An alternative scoring method would be to reflect one set of items or the other and then sum or average across all items. The results would be algebraically equivalent to this scoring method.

References

Aamodt, M. G., & Custer, H. (2006). Who can best catch a liar: A meta-analysis of individual differences in detecting deception. *Forensic Examiner*, **15**, 6–11.

Blair, J. P., Levine, T. R., & Shaw, A. J. (2010). Content in context improves deception detection accuracy. *Human Communication Research*, **36**, 420–439.

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., & The Global Deception Research Team. (2006). A world of lies. *Journal of Cross-Cultural Psychology*, **37**, 60–74.

Bond, G. D. (2008). Deception detection expertise. *Law and Human Behavior*, **32**, 339–351.

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

Ekman, P., & Friesen, W. V. (1969). Nonverbal leakage and clues to deception. *Psychiatry*, **32**, 88–106.

Exline, R. V., Thibaut, J., Hickey, C. B., & Gumpert, P. (1970). Visual interaction in relation to Machiavellianism and an unethical act. In R. Christie & F. L. Geis (Eds.), *Studies in Machiavellianism* (pp. 53–76). New York: Academic Press.

Frank, M. G., & Ekman, P. (2004). Appearing truthful generalizes across different deception situations. *Journal of Personality and Social Psychology*, **86**, 486–495.

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.

Goffman, E. (1956). The nature of deference and demeanor. *American Anthropologist*, **58**, 473–502.

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.

Levine, T. R. (2007). *MSU trivia game interviews.* Unpublished video tapes. East Lansing: Michigan State University.

Levine, T. R. (2010). A few transparent liars. *Communication Yearbook*, **34**, 40–61.

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.

Levine, T. R., Shaw, A., & Shulman, H. (2010). Increasing deception detection accuracy with strategic direct questioning. *Human Communication Research*, **36**, 216–231.

Miller, G. R., & Stiff, J. B. (1993). *Deceptive communication.* Newbury Park, CA: Sage.

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

O'Sullivan, M., & Ekman, P. (2004). The wizards of deception detection. In P. A. Granhag & L. Stromwell (Eds.), *The detection of deception in forensic contexts* (pp. 269–286). London: Cambridge University Press.

O'Sullivan, M., Frank, M. G., Hurley, C. M., & Tiwana, J. (2009). Police lie detection accuracy: The effects of lie scenario. *Law and Human Behavior*, **33**, 530–538.

Serota, K. B., Levine, T. R., & Boster, F. J. (2010). The prevalence of lying in America: Three studies of self-reported lies. *Human Communication Research*, **36**, 2–25.

Weinberger, S. (2010). Airport security: Intent to deceive? *Nature*, **465**, 412–415.

Zuckerman, M., DeFrank, R. S., Hall, J. A., Larrance, D. T., & Rosenthal, R. (1979). Facial and vocal cues of deception and honesty. *Journal of Experimental Social Psychology*, **15**, 378–396.

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.

El Comportamiento del Emisor: Las Diferencias Individuales en la Creencia del Emisor Tienen un Impacto Poderoso sobre los Juzgamientos en la Detección de la Decepción

El comportamiento del emisor es una diferencia individual en la creencia del emisor de un mensaje que es conceptualmente independiente de su honestidad actual. La investigación reciente sugiere que el comportamiento del emisor puede ser la sola fuente más influente de variación en los juzgamientos de detección de decepción. El comportamiento del emisor fue variado en 5 experimentos (N = 30, 113, 182, 30, 35) para crear condiciones de comparación igual entre comportamiento-veracidad y un desajuste entre comportamiento-veracidad. La inducción del comportamiento del emisor explicó cerca del 98% de la variación en la detección exacta. Tres estudios adicionales (N = 30, 113, 104) investigaron los perfiles de comportamiento de los emisores más o menos creíbles. Los resultados documentaron el impacto profundo de los efectos del emisor en la detección de la decepción y proveyeron de una explicación sobre el techo de baja precisión en estudios previos.

Palabras claves: mentir, decepción, diferencias individuales, exactitud, asesoramiento de la credibilidad

Das Auftreten des Senders: Individuelle Unterschiede bei der Beurteilung der Glaubwürdigkeit des Senders haben einen starken Einfluss auf die Wahrnehmung von Täuschungsversuchen

Das Auftreten des Senders stellt einen individuellen Unterschied der Glaubwürdigkeit des Botschaftssenders dar und ist konzeptionell unabhängig von der tatsächlichen Ehrlichkeit. Aktuelle Forschung legt nahe, dass das Auftreten des Senders die einflussreichste Quelle von Variation in der Beurteilung von Täuschungswahrnehmungen ist. Das Auftreten des Senders wurde in fünf Experimenten variiert (N = 30, 113, 182, 30, 35), um Konditionen mit stimmigen und unstimmigen Auftreten-Wahrhaftigkeit-Kombinationen zu schaffen. Das Auftreten des Senders erklärte bis zu 98% der Varianz bei der Wahrnehmungsgenauigkeit. In drei weiterführenden Studien (N = 30, 113, 104) untersuchten wir die Verhaltensprofile von mehr oder weniger glaubwürdigen Sendern. Die Ergebnisse verdeutlichen den starken Einfluss von Sendereffekten auf die Wahrnehmung von Täuschung und bieten zudem eine Erklärung für Deckeneffekte mit geringer Genauigkeit in anderen Studien.

Schlüsselbegriffe: Lügen, Täuschung, individuelle Unterschiede, Genauigkeit, Glaubwürdigkeitsmessung

Le comportement de l’émetteur : les différences individuelles dans la crédibilité de l’émetteur ont une influence puissante sur les jugements de détection du mensonge

Le comportement de l’émetteur est une différence individuelle dans la crédibilité des émetteurs de messages qui est conceptuellement indépendante de l’honnêteté réelle. Des études récentes laissent entendre que le comportement de l’émetteur pourrait être la source de variation la plus influente dans les jugements de détection du mensonge. Le comportement de l’émetteur a été altéré dans cinq expériences (N = 30, 113, 182, 30, 35) pour créer des conditions où le comportement correspondait à la véracité et d’autres où il n’y correspondait pas. L’induction basée sur le comportement de l’émetteur a expliqué jusqu’à 98 % de la variance dans l’exactitude de la détection. Trois études supplémentaires (N = 30, 113, 104) ont examiné les profils comportementaux d’émetteurs plus ou moins crédibles. Les résultats documentent l’impact puissant des effets d’émission dans la détection du mensonge et offrent une explication du bas niveau d’exactitude dans les résultats précédents.

Mots clés : mensonge, différences individuelles, exactitude, évaluation de la crédibilité

발송자 행위: 발송자의 신뢰에서의 개인적 차이들은 사기발견판단에 강력한 영향을 미친다.

발송자행위는 개념적으로 실제 정직한가와는 별개인 메시지 발송자의 신뢰도에서의 개인적 차이이다. 최근 연구는 발송자의 행위는 사기발견판단에서 가장 중요한 요소라는 것을 보여주고 있다. 행위-진실성 합치상황과 행위-진실성 불합치 상황을 만들기 위한 5 개 실험 (N=30, 113, 182, 30, 35)에서 발송자행위는 모두 다르게 나타났다. 발송자행위귀납은 발견의 정확성에서 변수의 98%를 설명하였다. 세가지 추가적 연구들 (N=30,113, 104)은 다소간 신뢰도가 있는 발송자들의 행위적 프로파일을 조사하였다. 결과들은 발송자의 강한 영량력은 사기발견에 효과적이었으며 이전 발견에서의 낮은 정확성을 설명하였다.