Personality and Social Psychology Review
2006, Vol. 10, No. 3, 214–234

Accuracy of Deception Judgments

Charles F. Bond, Jr.
Department of Psychology
Texas Christian University

Bella M. DePaulo
Department of Psychology
University of California at Santa Barbara

We analyze the accuracy of deception judgments, synthesizing research results from 206 documents and 24,483 judges. In relevant studies, people attempt to discriminate lies from truths in real time with no special aids or training. In these circumstances, people achieve an average of 54% correct lie–truth judgments, correctly classifying 47% of lies as deceptive and 61% of truths as nondeceptive. Relative to cross-judge differences in accuracy, mean lie–truth discrimination abilities are nontrivial, with a mean accuracy d of roughly .40. This produces an effect that is at roughly the 60th percentile in size, relative to others that have been meta-analyzed by social psychologists. Alternative indexes of lie–truth discrimination accuracy correlate highly with percentage correct, and rates of lie detection vary little from study to study. Our meta-analyses reveal that people are more accurate in judging audible than visible lies, that people appear deceptive when motivated to be believed, and that individuals regard their interaction partners as honest. We propose that people judge others' deceptions more harshly than their own and that this double standard in evaluating deceit can explain much of the accumulated literature.

Deception entered Western thought in a telling guise when the author of *Genesis* placed a serpent in the Garden of Eden. By lying, the serpent enticed Eve into committing the original sin. Thus deception was enshrined as the ultimate source of evil.

Lying has always posed a moral problem. Aristotle wrote that "falsehood is in itself mean and culpable;" St. Augustine believed that every lie is a sin; and Kant regarded truthfulness as an "unconditional duty which holds in all circumstances." Others take a more permissive stance. Aquinas countenanced lies told in the service of virtue, and Machiavelli extolled deceit in the service of self. For background on these ethical matters and a contemporary position, see Bok (1989).

Having been a moral issue for millenia, deception came also to be viewed as a legal challenge. Since Diogenes, many had suspected that lying was commonplace and could have pernicious influences on human affairs. The chore of truth finding fell to the legal system, and procedures for lie detection were devised. Over the centuries, authorities employed a number of unsavory means to extract legal "truths" (Trovillo, 1939). Modern sensibilities inspired some of the current techniques: religious oaths, cross-examinations, threats of incarceration. Technological developments have had an impact, too. The polygraph, the psychological stress evaluator, brain fingerprints, EEGs—these have been promoted for their ability to divine deception. Yet in the first decade of the 21st century, American jurisprudence entrusts lie detection to ordinary citizens. U.S. courts bar technological aids to lie detection and deception experts, too. Witnesses must appear in person before jurors who are the "sole judges" of the witnesses' believability. American jurors are instructed to judge the person's truthfulness by considering his or her "demeanor upon the witness stand" and "manner of testifying" (Judicial Committee on Model Jury Instructions for the Eighth Circuit, 2002, p. 53). According to an official view, this system of lay judgment solves the legal problem of deception because "lie detecting is what our juries do best" (Fisher, 1997, p. 575).

A moral problem for millenia and a legal problem for centuries, deception has more recently become a research problem. How successful are people at deceiving others? How likely are they to believe others' fibs? What accounts for liars' successes and failures? When and why are people duped? These questions are of moral and legal interest. The ethics of lying would be moot if people were rarely duped. Current legal prac-

We are grateful to Harris Cooper, Laura Muhlenbruck, Bob Rosenthal, and Jenny Tornqvist for help with this project. We also thank the many deception researchers who answered questions about their work.

Correspondence should be sent to Charles F. Bond, Jr., Department of Psychology, Box 298920, Texas Christian University, Fort Worth, TX 76129. E-mail: C.BOND@TCU.EDU

tices would be called into question if ordinary people could not spot deception when they saw it.

In this article, we summarize research on 4,435 individuals' attempts to dupe 24,483 others. We offer quantitative measures of deceptive success and identify conditions under which people are more and less gullible. As a background for our statistical synopses, we summarize some earlier characterizations of deception, sketch a new framework for understanding this subject, and consider earlier research reviews.

Characterizations of Deception

> "No mortal can keep a secret. If his lips are silent, he chatters with his finger-tips; betrayal oozes out of him at every pore." Freud (1905)

With this quotation, Ekman and Friesen (1969) opened a pioneering article on the psychology of deception. Where Freud had analyzed verbal slips to self-deception, Ekman and Friesen described nonverbal signs of individuals' attempts to deceive one another. These authors discussed lies that involve high stakes and strong emotion. In their view, liars face challenges. They must guard against nonverbal "leakage" of feelings they are trying to conceal and must hide their own affective reactions to the act of lying, such as guilt, anxiety, and shame. People find it especially difficult to lie in certain situations: when the possibility of deception is salient to both actor and target, when the target can focus on detecting deception without concern for his or her own behavior, and when the actor and target have antagonistic interests (the actor wishing to perpetrate deceit and the target to uncover it).

Ekman and Friesen (1969) offered a theory about the anatomical locus of nonverbal deception cues. They predict that people are most likely to show deception in the legs and feet, less likely to show it in the hands, and least likely to show deception in the face. These predictions followed from a communicative analysis: Relative to the face, the feet and legs have a weak sending capacity, generate little internal feedback, and occasion few reactions from others. Thus, people have more ability and motivation to control the face than the feet and legs. By this logic, people have intermediate ability and motivation to control the hands.

Thirty-two years later, Ekman (2001) emphasized the ambiguity of nonverbal deception cues. There being no foolproof sign of deceit, many inferences of deception are mistaken. In trying to spot lies, people must avoid untoward influences of their own suspicions as well as misinterpretations of others' idiosyncrasies. Ekman attributes failures at lie detection to many factors: poor evolutionary preparation, socialization to overlook lies, the psychological benefits of trust, and inadequate feedback from errors.

Ekman's work has been influential. It has encouraged nonverbal analyses that aim to expose deceit. Inspired by Ekman's early work, Miller and Stiff (1993) enumerated cues to deception and cues to judgments of deception, then attributed failures at spotting deception to differences in the two sets of cues. Pursuing Ekman's emphasis on high-stakes deceit, forensic psychologist Vrij (2000) discussed the implications of experimental findings for professional lie catchers.

Buller and Burgoon (1996) proposed a theory for face-to-face deceptive interactions. To dupe others, people must craft a verbal deception, bolster it with ancillary strategic messages, and suppress discrediting behaviors. Meanwhile, the targets of face-to-face deceit must manage behavioral signs of suspicion. Buller and Burgoon (1996) traced the unfolding of deceptive exchanges over time. Theoretically, receivers are more likely to perceive a person as truthful if they are interacting with that person rather than seeing the person on videotape. Theoretically, deceivers should be more likely to engage in strategic activity and less likely to engage in nonstrategic activity in interactive contexts. In interactive contexts, deceivers react to signs of suspicion, and targets react to indications that their suspicions have been surmised.

Critical of Buller and Burgoon's (1996) approach (DePaulo, Ansfield, & Bell, 1996), DePaulo and colleagues favored a self-presentational perspective on deception (DePaulo, 1992; DePaulo, Kashy, Kirkendol, Wyer, & Epstein, 1996; DePaulo et al., 2003). In this view, lying is a part of everyday life. People tell lies to avoid embarrassment and make positive impressions. They fib on the spur of the moment without compunction, telling polite lies of little consequence. Some everyday lies are scripted and require less cognitive effort than meticulously truthful statements. Occasionally, people tell lies to hide transgressions. Most of these serious lies involve a self-presentational stake: the liar's reputation. In this view, the signs of deception are subtle, and social norms encourage people to accept others' representations at face value.

A Double Standard

Having reviewed some earlier characterizations of deceit, we offer a new framework for understanding this subject. We believe that there is a double standard in evaluating deception.

Our framework begins by noting that people regard truth telling as unexceptional. They accept most statements at face value, rarely inquiring into the authenticity of what they hear. People come to regard an assertion as truthful only after entertaining the possibility that it was deceptive. Then they see truthfulness as a virtue. People are proud of themselves for speaking the truth. People who are told the truth praise truth tellers, and psychologists praise them, too. No doubt, there are limits to the morality of truthfulness. Truths are seen as most virtuous when they oppose the truth teller's interest. Occasionally, people volunteer truthful observations that hurt others, and these are ethically dubious. In most cases, however, truth telling is nonproblematic. Problems arise not from truth telling but from deception.

There are two perspectives on deception. One is the perspective that people hold when they themselves are lying; a second is the perspective they bring to others' lies (Gordon & Miller, 2000). As deceivers, people are practical. They accommodate perceived needs by lying. Of course, deceivers rarely regard their own falsehoods as lies but as something more innocuous. People may lie in the interest of impression management (DePaulo et al., 2003) or for more tangible ends. They exaggerate, minimize, and omit. They give misleading answers to questions. Regarding half-truths and self-editing as necessities of social life, deceivers see deception as similar to these sanctioned practices. Animated by momentary exigencies, offered in passing, lies occasion little anxiety, guilt, or shame (DePaulo, Kashy, et al., 1996). They are easy to rationalize. Yes, deception may demand construction of a convincing line and enactment of appropriate demeanor. Most strategic communications do. To the liar, there is nothing exceptional about lying.

If pragmatic about their own deceptions, people become moralistic when they consider others' lies (Saxe, 1991). Then deception is wrong and reflects negatively on the deceiver. Indeed, people view duplicity as one of the gravest moral failings. In their ratings of 555 personality trait terms, college students rate as least desirable the trait of being a liar (Anderson, 1968). Social logic assumes that honest people always act honestly (Reeder, 1993); thus, to label a statement a lie is to imply that the person who made that statement is a liar (O'Sullivan, 2003). This is a serious accusation. People have a prescriptive stereotype of the liar—stricken with shame, wracked by the threat of exposure, liars leak signs of their inner torment. They fidget, avoid eye contact, and can scarcely bring themselves to speak—a worldwide stereotype holds (Global Deception Research Team, 2006). The stereotypic reasons for lying are nefarious, too—terrorists lying to further their murderous plots, charlatans scheming to bilk the innocent, husbands cheating on their faithful wives. As old as the Garden of Eden, this moralistic perspective on deceit underlies current psychological thought.

Let us sketch a few implications of the double standard in evaluating deception. People hold a stereotype of the liar—as tormented, anxious, and conscience stricken. Perceivers draw on this stereotype when considering a target's veracity. Targets who most resemble the stereotype are most likely to be regarded as liars; those who least resemble it are most likely to be believed. Factors that influence a person's likelihood of appearing tormented, anxious, or conscience stricken should affect the person's judged truthfulness. One such factor would, we suspect, be the stakes surrounding a speaker's credibility. Most lies are little. When telling white lies of the sort often studied by researchers, people have no reason to appear tormented. Thus, they should often be judged truthful. Occasionally, the stakes of being believed are big. When facing huge stakes, people who ruminate over their credibility may come to match the liar stereotype. Then they would be judged deceptive, even if they were telling the truth.

In this article, we consider veracity judgments in light of the double standard for evaluating deception. We do not confine attention to momentous lies or evil deceit of the sort most would associate with others' deception. Rather, we consider all falsehoods that have been studied and hope to use the accumulated literature to learn about people's successes in engineering various forms of deception. We will credit people for their successes at perpetrating deceit and note some unintended consequences of observers' moralistic stance.

Research on Detection Accuracy

To understand deception, researchers conduct experiments. They arrange for people to lie and tell the truth, and for others to judge the veracity of the resulting statements. For convenience, we call the people who lie in these experiments *senders*, the truthful and deceptive statements *messages*, and the people who judge these messages *receivers*. We are interested in receivers' accuracy in judging senders' veracity. We do not review all attempts at lie detection. Rather, we confine attention to receivers who must judge deceit without the aid of polygraphs, fMRIs, or other physiological devices—receivers who judge deception from a brief encounter with an unfamiliar sender in real time. These deception judgments are based on verbal content and the liar's behavior. Here we review earlier summaries of this research.

Often, lie detection abilities are expressed on a familiar scale: percentage correct. In relevant experiments, receivers classify messages as either lies or truths; hence, across messages, the percentage of messages a receiver correctly classifies can be used as an index of his or her detection ability. Ordinarily, half of the messages a receiver encounters are truths, and half are lies; hence, by guessing a receiver could expect to achieve 50% correct classifications.

Kraut (1980) offered a statistical summary of results from 10 such experiments. Finding a mean accuracy rate of 57%, Kraut concluded that "the accuracy of human lie detectors is low." In a summary of 39 studies published after 1980, Vrij (2000) replicated Kraut's finding, discovering that receivers of more recent research achieve an average of 56.6% accuracy.

Along with narrative reviews of the research literature, these statistical summaries have inspired a consensus—"it is considered virtually axiomatic … that individuals are at best inaccurate at deception detection" (Hubbell, Mitchell, & Gee, 2001, p. 115).

Although it may be "virtually axiomatic" that people are poor at detecting deception, we are reluctant to accept this conclusion on the basis of existing work. We agree that in 50 (or so) pertinent studies people achieved 50% to 60% correct when classifying messages as lies or truths. However, meta-analyses of percentage correct omitted evidence relevant to ascertaining the accuracy of deception judgments. In the omitted experiments, receivers rated the veracity of lies and truths on multipoint rating scales. Accuracy was not gauged in terms of percentage correct but as a difference between the rated veracity of truths versus the rated veracity of lies.

Three statistical summaries of lie detection accuracy have incorporated rated-veracity results. They quantify the degree to which lies can be discriminated from truths by a standardized mean difference (d): the mean difference between obtained and chance accuracy in a study divided by a standard deviation from that study. Applying this metric to the results of 16 early studies, DePaulo, Zuckerman, and Rosenthal (1980) calculated a median d of .86 standard deviations. Twenty years later, Mattson, Allen, Ryan, and Miller (2000) found an average difference between the judged veracity of lies and truths of $d = 1.07$ standard deviations in 7 studies of organizational deception. Assessing the accuracy of deception judgments in various media, Zuckerman, DePaulo, and Rosenthal (1981) found that receivers who have access to speech regard lies as less credible than truths with a mean $d = 1.14$.

How strong are the levels of lie detection found in these rated-veracity reviews? To answer this question, it may be helpful to consider results found in other lines of research. From a large-scale compilation, Richard, Bond, and Stokes-Zoota (2003) developed empirical guidelines for evaluating effect sizes. These scholars described a d of .20 as small, a d of .40 as medium, and a d of .60 as large because these values would be larger than the average standardized mean differences found in 30%, 50%, and 75% of 474 social psychological research literatures the scholars reviewed. Compared with these reference values, people would seem to have a strong ability to detect deception. The median d of .86 standard deviations found by DePaulo et al. (1980) would place lie detection accuracy at roughly the 85th percentile in size, relative to 474 social psychological effects (Richard et al., 2003). The ability to detect audible lies (mean $d = 1.14$ standard deviations; Zuckerman et al., 1981) is even better, ranking at the 95th percentile of 474 social psychological effects.

While amassing evidence on receivers' accuracy in discriminating lies from truths, scholars have been interested in a more general judgmental tendency—a bias to perceive messages as truthful. By virtue of the bias, truthful messages are more often detected than deceptive messages. Summarizing 15 studies, Zuckerman et al. (1981) express this accuracy difference in standard deviation units and find a mean $d =$ .86. Vrij (2000) summarizes 9 percentage-correct studies to find a strong truth bias—a mean of 61.5% truth judgments, 67% accuracy for truths, and 44% accuracy for lies.

This Review

Given the moral and legal significance of deception, it is important to know how often people are duped. Although previous work provides some hints about people's success in deceiving others, the work has limitations. The largest review to date is based on 39 research studies. Here we summarize evidence from 206 studies. Some of the previous reviews express the accuracy of deception judgments as a standardized mean difference, whereas others gauge accuracy in terms of percentage correct. Each of these measures has limitations. Standardized mean differences can be hard to interpret (Bond, Wiitala, & Richard, 2003), and meta-analyses of percentage correct cannot include results on rating scale judgments of deception.

Here we assess the accuracy of deception judgments in terms of percentage correct, the standardized mean difference, and with some indexes that statisticians favor—the log odds ratio and d' (Swets, 1996). Perhaps the pattern of results across various measures of accuracy can help resolve a tension in earlier meta-analytic results—between the strong detection abilities implied by standardized results and an "axiom" of inaccurate lie detection in percentage correct (Hubbell et al., 2001).

Some have thought that detection performances vary only slightly across situations (Kalbfleisch, 1990), whereas others have concluded that performance variance across situations is considerable (Miller & Stiff, 1993). Here we provide the first test to date of the possibility that there is no variance in detection performances across situations. Assuming that there is such variance, we provide the first estimates to date of the magnitude of these situational differences. We also have the opportunity to document the impact of various factors on the accuracy of deception judgments, like the medium in which deception is attempted, the liar's motivation, and the judge's expertise. The evidence may have implications for theories of deception, including our double standard framework.

Method

Literature Search Procedures

To locate relevant studies, we conducted computer-based searches of *Psychological Abstracts*, *PsycInfo*, *PsycLit*, *Communication Abstracts, Dissertation Abstracts International*, *WorldCat*, and *Yahoo* through August of 2005 using the keywords *deception*, *deceit*, and *lie detection*; searched the *Social Sciences Citation Index* for papers that cited key references, examined reference lists from previous reviews (DePaulo, Stone, & Lassiter, 1985; Zuckerman et al., 1981); and reviewed the references cited in more than 300 articles on the communication of deception from our personal files plus all references cited in every article we found. We sent letters requesting papers to scholars who had published relevant articles.

Criteria for Inclusion and Exclusion of Studies

Our goal was to summarize all English-language reports of original research on the accuracy of judgments of lies and truths available to us prior to September 2005. To be included in this review, a document had to report a measure of accuracy in discriminating lies from truths.

We excluded studies in which individuals judged only lies and those in which individuals judged only truths. We excluded studies in which judges received experimental training or instructions about how to detect deception, studies in which judges received attention-focusing instructions, studies in which senders and receivers knew one another prior to the study, and studies in which individuals could incorporate into their judgments systematic aids to lie detection (e.g., polygraph records, Criterion-Based Content Analysis, or behavior codings from repeated viewings of a videotape). We excluded reports that were not in English, judgments for lies and truths told by senders who were less than 17 years old, and judgments made by receivers who were less than 17. We excluded reports in which senders role-played an imagined person in an imagined situation. We also excluded all results on implicit deception judgments (implicit judgments having recently been meta-analyzed by DePaulo et al., 2003), and on judgments of affect (even affects that people were trying to conceal). We uncovered 206 documents that satisfied our inclusion criteria. For a listing of these documents, see Appendix A, which is available online at: http://www.leaonline.com/doi/pdf/10.1207/s15327957/pspr1003_2A.

Identifying Independent Samples

Research studies in this literature exhibit two forms of interdependence: sender interdependence and receiver interdependence. Senders are interdependent when the lies and truths told by a given sample of senders are shown to multiple samples of judges. Receivers are interdependent when researchers report multiple measures of lie–truth accuracy for a given sample of judges. The unit of aggregation in the current meta-analysis is the receiver sample. The following primary analyses extract one measure of lie–truth discrimination accuracy from each independent sample of judges—even in those cases in which several samples are judging the same lies and truths. For these analyses, our data set consists of 384 independent samples. To assess the impact of moderator variables, we disaggregated receiver samples to reflect within-receiver experimental manipulations.

Variables Coded From Each Report

From each report, we sought information about the following variables: (a) number of senders, (b) number of receivers, (c) percentage correct, (d) percentage truth, (e) an accuracy standardized mean difference, (f) sender motivation, (g) receiver motivation, (h) sender preparation, (i) sender interaction, (j) receiver expertise, (k) judgment medium, and (l) baseline exposure. For our coding of these variables in each of 384 receiver samples, see Appendix B online at: http://www.leaonline.com/doi/pdf/10.1207/ s15327957/pspr1003_2B.

Let us explain these variables. The number of senders and number of receivers were coded from each document. From each document that reported results on dichotomous lie-or-truth classifications, we noted percentage correct—more precisely, the unweighted average of the percentage of truthful messages correctly classified and the percentage of deceptive messages correctly classified. Of our 384 receiver samples, 343 judged 50% lies and 50% truths. In these cases, the unweighted average was the overall percentage correct. Whenever authors reported the overall percentage of messages classified as truthful, this *percentage truth* judgments was coded. From each document that reported results on rating-scale veracity judgments, we noted an accuracy standardized mean difference—defining d as the mean veracity rating of truths minus the mean veracity rating of lies divided by a standard deviation. As Kalbfleisch (1990) noted, deception researchers' reporting of standard deviations poses challenges for meta-analysts. Whenever possible, we used as our standard deviation a pooled within-message standard deviation across receivers. In such cases, we would note the variance across receivers in judgments of the veracity of truthful messages and the variance across receivers in judgments of the veracity of deceptive messages, before taking the square root of the average of these two variances. When necessary, we used other standard deviations—for example, the standard deviation across receivers in the difference between the mean rated veracity of truths and the mean rated veracity of lies.

The other variables of interest to us are categorical. People can try to detect lies over various media. Here we coded *deception medium* by noting whether a given sample of receivers was trying to detect lies over a video medium, an audio medium, an audiovisual medium, or some other medium. We coded *sender motivation* by noting whether participants had any special motivation to succeed at deception. Our coding of *sender preparation* reflected whether the senders in a study had any time to prepare their lies and truths. We coded whether or not receivers got a baseline exposure to the sender before making deception judgments.

In some studies, senders are interacting with others as they lie and tell the truth; in other studies, they are not. For purposes of coding *sender interaction*, we regarded senders as not interacting if when lying they were alone or in the presence of a passive observer. We deemed all other senders to be interacting and noted whether or not the interaction partner was the receiver (e.g., the person who was judging deception). Most of the receivers in this literature are college students. Others are people whose occupations are thought to give them special expertise at lie detection. We noted this variable of *receiver expertise*.

We coded the status of the report as published or unpublished. In some instances, the same data are reported in two places—say, a dissertation and a journal article. In such cases, we have listed the more accessible report in the References section. Occasionally, results from a given study are more fully reported in one document than another. Then we used the more complete reporting even if it was from the less accessible document.

Reliability of Coding

For a reliability check, the two authors independently coded 24 of the documents (Appendix A). These were selected at random, subject to the restriction that no individual appear as an author on more than two documents. The 24 documents we selected in this manner contribute 46 independent receiver samples to our meta-analysis, and it is on these 46 receiver samples that reliability data are available. The following quantitative variables were checked: number of senders, number of receivers, percentage correct, percentage truth, and accuracy d. Reliabilities on these variables were uniformly high; lowest Pearson's $r = .894$ for 10 accuracy ds. We also checked coding of the following categorical variables: sender motivation, receiver motivation, sender preparation, sender interaction, judgment medium, and baseline exposure. For the percentage agreement on each of these variables, see Table 1.

Results

Characteristics of the Literature

We found 206 documents that satisfied our criteria—133 that were published and 73 that were unpublished. The earliest document was dated 1941, and the latest was published in 2005. Half of these documents were dated 1994 or earlier.

The documents reported results on 24,483 receivers' deception judgments of 6,651 messages offered by 4,435 senders. There were 177 independent samples of senders and 384 independent samples of receivers. One hundred ten of the sender samples were judged by only a single receiver sample; at the other extreme, one sample of senders was judged by 13 independent receiver samples.

In 277 receiver samples, participants classified messages as lies or truths; in 92 samples, they judged messages on multipoint rating scales; and in 15 samples, receivers made lie-or-truth classifications as well as multipoint ratings. For some other characteristics of this literature, see Table 1. In a typical research study, 41 receivers made judgments of 16 messages—one message offered by each of 16 senders. The typical message lasted 52 sec. In most cases, the judgment medium was audiovisual, and receivers had no baseline exposure to the sender. Although about 55% of the sender samples had no particular motivation to succeed when lying, more than 40% were motivated. Receivers were rarely motivated; barely 12% of the receiver samples had any special incentive to succeed at lie detection. In a little more than half of the samples, receivers were judging senders who had had time to prepare their lies; in about 65% of the samples, receivers judged senders who were interacting as they lied. Although only 12% of the receiver samples could claim any occupational expertise in detecting deception, this was nonetheless 2,842 experts.

Percentage Correct

In 292 samples, receivers classified messages as lies or truths. From each such sample, we noted the mean percentage correct lie–truth classifications. These are shown on the right side of Figure 1 as a stem-and-leaf display. As can determined from the display, more than three fourths of these means are greater than 50% and less than one in seven is greater than 60%. Across all 292 samples, the unweighted mean percentage correct lie–truth classifications is 53.98%. The highest mean percentage correct attained in any sample is 73%, and the lowest is 31%. Means at the first, second, and third quartile are 50.07%, 53.90%, and 58.00%.

Further insight into lie–truth discrimination abilities can be gleaned from Figure 2, which displays the mean percentage correct lie–truth classifications in a study as a function of the total number of judgments on which the mean was based. The latter was determined

Table 1. *Characteristics of the Research Literature*

	Quantitative Variables				
Variable	Minimum	Maximum	Mean	Median	*s*
Number of Senders	1	200	22.45	16.00	22.63
Number of Receivers	1	816	63.65	41.50	70.56
Messages per Receiver	1	416	31.89	16.00	44.50
Message Duration (Sec)	2	1200	110.63	52.00	173.16

	Categorical Variables	
Variable	No (%) of Receiver Samples	Percent Coding Agreement
Deception Medium		91.3%
Video	47 (12.2%)	
Audio	42 (10.9%)	
Audiovisual	262 (67.4%)	
Other	22 (4.9%)	
Within-Receiver Manipulation	11 (4.4%)	
Sender Motivation		89.5%
No Motivation	214 (55.7%)	
Motivation	153 (39.8%)	
Within-Receiver Manipulation	17 (4.4%)	
Sender Preparation Time		81.1%
None	196 (51.0%)	
Some	165 (43.0%)	
Within-Receiver Manipulation	23 (6.0%)	
Baseline Exposure		91.3%
No Exposure	360 (93.7%)	
Exposure	20 (5.2%)	
Within-Receiver Manipulation	4 (1.1%)	
Sender Interaction		100%
None	127 (33.1%)	
Interaction With Receiver	33 (8.6%)	
Interaction With Another	224 (58.3%)	
Receiver Expertise		100%
Not Expert	338 (88.0%)	
Expert	46 (12.0%)	

by multiplying the number of receivers in a sample by the number of judgments each receiver rendered. Note, for example, the rightmost point in the plot. This represents the mean lie–truth discrimination accuracy of 54.30% observed by DePaulo and Pfeiffer (1986) in 10,304 dichotomous lie–truth judgments (64 judgments made by each of 161 receivers).

Figure 2 exhibits a funnel pattern (Light, Singer, & Willett, 1994) with high variability among means based on small numbers of judgments and low variability among means based on large numbers of judgments. This pattern suggests that the studies are estimating a common value and that small sample sizes account for much of the variability toward the left of the plot.

A formal analysis of between-study differences begins by noting that the observed standard deviation in mean percentage correct is only 6.11% (that is, variance = 37.33%). Statistically, results would vary some from study to study merely by virtue of different investigators examining different receivers. Random-effects techniques can be used to separate between-study variance due to sampling variability from true variance (Hedges & Vevea, 1998). Using a weighted method of moments technique, we infer that receiver sampling error accounts for 45.29% of the observed between-study variance in mean percentage correct, and that the true standard deviation across studies in mean percentage correct is only 4.52%.

For other analyses of mean percentage correct, we used procedures outlined by Bond et al. (2003). These require an estimate of the standard deviation in percentage correct in each study. Whenever a standard deviation was reported (or could be calculated), we used it. Otherwise, we imputed the standard deviation across the receivers in a sample from the binomial distribution, using the mean sample percentage correct as well as the number of judgments made by each receiver in that sample.

Mean percentage truth judgments		*Mean percentage correct lie/truth judgments*
(*k* = 207)		(*k* = 292)
Leaves	***Stem***	***Leaves***
0	**9**	
	8	
666	**8**	
	8	
322	**8**	
110	**8**	
98	**7**	
76	**7**	
5544	**7**	
2	**7**	3
110000	**7**	11
98888888	**6**	
7777	**6**	667777
55555554	**6**	444455
333222	**6**	2222222333
1111000000000	**6**	0000000000000111111
9999999988888888	**5**	8888888888888888899999999999999
7777777777777666666	**5**	666666667777777777777777
555555544444444444	**5**	444444444444444444444444455555555555555555555555
3333333333333332222222222222	**5**	222222222222222233333333333333333333
1111100000000	**5**	000000000000000000000111111111111111111111
99999988888	**4**	888888888888899999999999
77776666666666666666666	**4**	66666667777777777777
5544	**4**	4455555
33222	**4**	2222
	4	0011
9	**3**	99
	3	7
55	**3**	5
	3	
10	**3**	1
9	**2**	
77	**2**	

Figure 1. Stem-and-leaf plots.

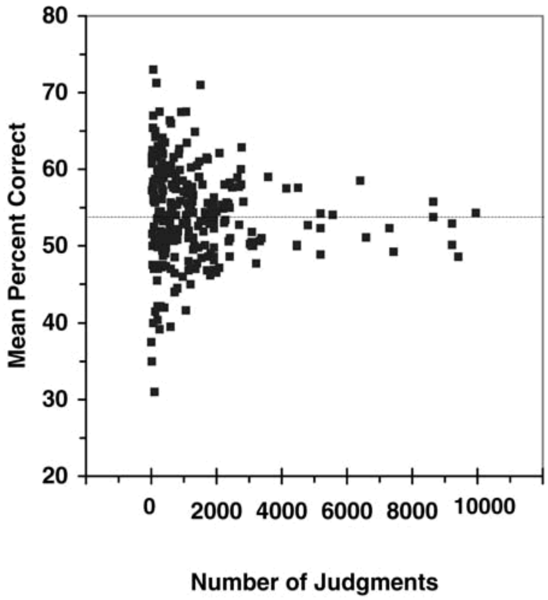

Figure 2. Mean percent correct by number of judgments.

These weighted techniques reveal a mean of 53.46% correct lie–truth classifications; 95% confidence interval = 53.31% to 53.59%. This mean is significantly greater than 50%, $t(7{,}994) = 39.78$, $p < .0001$. Between-study variability (though small in size) is greater than would be expected by chance, $F_w(283, 3658) = 12.61$, $p < .0001$.

Standardized Mean Differences

Having found that dichotomous lie-or-truth classifications are correct slightly more than half of the time, we next wished to gauge receivers' ability to distinguish lies from truths on multipoint rating scales. In relevant studies, accuracy is operationalized as the mean honesty rating of truthful messages minus the mean honesty rating of deceptive messages. Because different rating scales are used in different studies, it is necessary to standardize these results before summarizing them. To do so, we divide the mean difference in a study between the rated honesty of truths and lies by a standard deviation from that study. Earlier meta-analyses gave us reason to imagine that rating-scale lie–truth discrimination might be sizeable, yielding mean differences in the range of .86 standard deviations (DePaulo et al., 1980) or 1.14 standard deviations (Zuckerman et al., 1981).

We found 107 samples of receivers who rated deceptive and truthful messages on multipoint scales. For each of these samples, we computed a standardized difference between means (d). The unweighted mean d was .35 ($s = .47$). The ds at the first, second, and third quartile were .09, .31, and .67. By fixed-effects methods (Lipsey & Wilson, 2001), the weighted mean d for lie–truth discrimination is .34; 95% confidence interval = .31 to .38. There is statistically significant heterogeneity in the size of these Cohen's ds, $Q(106) = 458.74$, $p < .01$. Receiver sampling error accounts for 21.92% of the observed variance in effect sizes, and the true standard deviation in these standardized mean differences is .37. It is noteworthy that the level of lie–truth discrimination we find in 107 studies of rated veracity (mean $d = .35$) is less than half as large as the levels reported in earlier rating reviews (where ds exceeded .85).

Existing summaries led us to suspect that lies might be better discriminated from truths when the discrimination was attempted on multipoint rating scales rather than with dichotomous classifications. To assess this suspicion, we also computed a standardized mean difference for each study in which participants made lie-or-truth classifications. In such cases, the relevant measure is the mean percentage of truthful messages classified as truths minus the mean percentage of deceptive messages classified as truths divided by a standard deviation.

The dichotomous standardized mean differences yielded a weighted mean of .42 in 216 samples from which they could be computed. Values at the first, second, and third quartile were .02, .50, and 1.04. For 61 other samples, no standard deviation in percentage correct lie–truth classifications was reported. There we used the binomial distribution to impute a within-message standard deviation across receivers and found a weighted mean d of .40.

As these computations indicate, the standardized mean difference in the perceived truthfulness of truths and lies is smaller when receivers use rating scales, rather than when they make lie-or-truth classifications, weighted mean ds = .34 versus 41; for the difference, $Q(1) = 8.86$, $p < .05$. Combining together lie–truth discrimination results from all 384 receiver samples, we find weighted and unweighted mean ds of .39 and .49, respectively. The median d is .39. Standardized mean differences can be converted to Pearson product–moment correlation coefficients. If we convert each d to an r and cumulate the latter in the usual way, we find an unweighted mean accuracy $r = .19$ and r corresponding to the weighted Fisher's $Z = .21$.

Here lie–truth discrimination abilities produce a weighted mean d of approximately .40. This is considerably smaller than the ds of .86, 1.07, and 1.14 reported in earlier rated-veracity reviews. Even so, the ability to discriminate lies from truths at this level should not be dismissed. Many widely cited effects in social psychology are smaller than this one. Indeed, our d of .39 (or r of .21) would rank above the 60th percentile in size, relative to 474 social psychological effects compiled by Richard et al. (2003).

Percentage Judged True

Deception judgments can have large consequences whether or not they are correct. Thus, it is important to understand factors that may bias the judgments in one direction or another. Vrij (2000) reviewed evidence for a truth bias—receivers' tendency to err in the direction of judging messages as true.

Researchers reported the percentage of messages receivers classified as true in 207 receiver samples. These are displayed on left side of the stem-and-leaf plot in Figure 1. As Figure 1 shows, the percentage of truth classifications is higher than the percentage of correct classifications, and the percentage of truth classifications is more variable. Percentage truth classifications show an unweighted mean of 56.86% and weighted mean of 55.23%. Each of these values is significantly greater than 50%; for the weighted mean, $t(6,914) = 46.85$, $p < .0001$. The 95% confidence interval for the weighted mean percentage judged true extends from 54.99% to 55.46%, and the true standard deviation across studies in this percentage is 8.13.

Senders succeed in conveying more honesty than dishonesty in these studies. However, the bias thus introduced into receivers' judgments (of roughly 56% truth judgments) is smaller than the 61% truth judgments reported in a tabulation of 9 studies (Vrij, 2000). Across studies, there is no relationship between the percentage of truth judgments receivers rendered and the percentage of correct lie–truth classifications they achieved, $r = -.008$.

Stimulus Accuracy

Because an overall accuracy score is computed by averaging the percentage of correct classifications of truthful messages with the percentage of correct classifications of deceptive messages, it may seem informative to analyze separately the two component scores. We regard these two scores as indexes of stimulus accuracy for truthful messages and stimulus accuracy for deceptive messages, respectively.

In 207 receiver samples, percentage accuracy rates could be determined for truthful messages and deceptive messages separately. These are the same 207 samples used in our previous tabulation of the truth bias. Unweighted analyses reveal that people correctly classify 61.34% of truthful messages as truthful and 47.55% of deceptive messages as deceptive.

There is variability from study to study in the percentage correct classification of deceptive messages as well as truthful messages (each $s = 12.51\%$). The greater the percentage of lies in a study that are correctly classified, the lower is the percentage of truths in that study correctly classified; for the cross-study relationship, $r = -.53$, $p < .0001$. Across studies, accuracy at detecting lies shares little variance with accuracy at detecting truths. Any shared accuracy variance is overwhelmed by cross-study differences in suspicion. Thus, cross-study differences result largely from differences in response threshold, rather than differences in discrimination ability.

Response Accuracy

The questions of whether people can identify truths as truths and lies as lies, and of differential rates of success, are important ones. But they tell only part of the story about accuracy at detecting deception. Left unanswered are two parallel questions. Given that a person has judged a statement to be truthful, what is the likelihood that the statement was in fact truthful? And, given that a person has judged a statement to be a lie, what is the likelihood that it was actually a lie? To address these questions, we determined the response accuracy of a receiver's truth judgments and the receiver's lie judgments—defining them as the percentage of truth (and of lie) judgments that were correct. Recognizing that response accuracy scores could depend heavily on the baseline percentages of truthful and deceptive messages judged, we restricted our analyses of these measures to receivers who judged an equal number of deceptive and truthful messages. Unweighted means on the relevant 187 samples indicate that judgments of truthfulness are less likely to be accurate than are judgments of deceptiveness, unweighted means = 54.12% versus 55.84%, $t(187) = 5.12$, $p < .01$. There are cross-study differences in the response accuracy of lie and truth judgments ($s = 7.75$ and 5.84, respectively). Interestingly, the greater the response accuracy of truth judgments in a study, the greater is the response accuracy of lie judgments in that study, and this relationship is strong, $r = .80$, $p < .0001$.

Other Accuracy Measures

All the measures of accuracy that we have considered so far have limitations. Stimulus accuracy measures can be inappropriately affected by variations in judgmental bias, and response accuracy measures can be artifactually affected by variations in deception base rate. In light of these limitations, we analyzed this research literature with several alterative measures of lie–truth discrimination accuracy—including the log odds ratio and d'. These measures have a theoretical advantage over percentage correct, as they are statistically independent of variations in judgmental bias and base rate. We had imagined that these alternative measures might provide distinctive information about people's average ability to detect lies and give us new insights into cross-study differences in lie detection. They did not.

For one set of analyses, we used methods described in Fleiss (1994) to compute a detection accuracy odds ratio. This was a ratio of the odds that a truthful message was judged to be the truth (rather than a lie) divided by the odds that a deceptive message was judged to be the truth (rather than a lie). Aggregating log odds ratios across the 207 samples in this literature for which requisite data are available, a backtransformation of the mean indicates that the odds of judging a truthful message as the truth is 1.46 times as great as the odds of judging a deceptive message to be the truth (95% CI = 1.41–1.51). The back-transformed mean odds of truth detection is 1.65, and the corresponding mean for lie detection is .91. These imply means of 62.30% and 47.53% correct judgments to truthful messages and deceptive messages, respectively—quite close to the stimulus accuracy means of 61.34% and 47.55% directly computable from these samples.

Encouraged by signal detection theorists (e.g., Swets, 1996), we used a binormal method to calculate d' from each of 207 receiver samples. Here d' represents a mean difference in apparent honesty between deceptive and truthful messages. The mean d' in these studies is .24; the median is .22. Although a given d' can correspond to a number of different percentage correct lie–truth classifications, the maximum percentage correct would occur if the percentage correct judgments of deceptive messages equaled the percentage correct judgments of truthful messages (Walter, 2002). For the mean d' of .24 in this literature, this maximum is 54.79%—quite close to the mean of 54.45% correct directly computable from these samples.

Calculations with the odds ratio and d' corroborate the general conclusion we reached from analyzing percentage correct—that in the typical research setting lies are discriminated from truths at levels that are slightly better than would be attained by flipping a coin. To determine whether these alternative accuracy measures might give us distinctive information about cross-study differences in lie–truth discrimination, we computed some correlation coefficients across the relevant 207 receiver samples. Results reveal that the three accuracy measures we have been discussing are very highly intercorrelated. As an index of cross-study accuracy differences, percentage correct is virtually interchangeable with the log odds ratio (r = .979) and d' (r = .988). The latter two measures are barely distinguishable, r = .999. These results should be heartening to the many researchers who have been measuring lie detection accuracy as percentage correct.

Determinants of Accuracy

Thus far, our analysis indicates that individuals have some ability to detect deception. On the average, judges achieve about 54% lie–truth discrimination accuracy. As a percentage, lie–truth discrimination abilities seem poor, but when scaled by cross-judge standard deviations, these abilities appear nonnegligible. These are typical results over a variety of receiver samples, sender samples, deception media, types of lies, and contexts. Perhaps under certain conditions judges show high percentage lie–truth discrimination rates; perhaps under other conditions, they show trivial standardized discrimination performances. To assess these possibilities, we now examine various subsets of the research literature on deception judgments. We hope to determine how deception judgments are influenced by six factors: (a) deception medium, (b) motivation, (c) preparation, (d) baseline exposure, (e) interaction, and (f) receiver expertise.

Each of these factors will be assessed in its impact on three different indexes of judgment: (a) percentage truth classifications, (b) percentage correct lie–truth classifications, and (c) a standardized difference between the perceived veracity of truths and the perceived veracity of lies. The first two indexes were coded from studies in which individuals made dichotomous lie–truth classifications and were analyzed with the raw techniques of Bond et al. (2003). The third index, which included results from both lie-or-truth classifications and veracity ratings, was analyzed with standardized fixed effects techniques (Lipsey & Wilson, 2001).

To infer the effects of each factor, we consider three forms of evidence: within-study comparisons, between-study comparisons, and statistically adjusted comparisons. We aggregate within-study comparisons for each moderator variable that has been examined within studies. Summaries of relevant experiments provide us with controlled evidence of the impact of moderator variables. Unfortunately, the number of experiments that manipulate a given factor is limited, as is the range of conditions under which it has been examined. Thus, we also assess effects using between-study comparisons. We assess the effect of a person's motivation to lie, for instance, from a comparison of effect sizes in studies in which participants were motivated to lie with effect sizes in studies in which they were not motivated. Although we can base between-study comparisons on impressive amounts of data, the studies at one level of a moderator variable may differ in any number of ways from the studies at another level. In light of these potential confounds, we also make statistically adjusted comparisons. They gauge the impact of a given moderator variable from a multiple regression analysis that adjusts for the impact of other variables. In particular, our statistically adjusted comparisons of percentage truth classifications and percentage correct lie–truth classifications document the partial effect of a given moderator variable from an inverse variance weighted multiple regression equation that includes as regressors the six factors enumerated previously (deception medium, motivation, preparation, baseline exposure, interaction, and receiver expertise), as well as a control variable indicating whether or not messages were edited prior to presentation. Our statistically adjusted comparisons of ds reflect results from an inverse variance weighted multiple regression equation that includes the seven regressors just mentioned, as well as an eighth variable that indicates whether deception judgments were rendered as lie-or-truth classifications or on multipoint rating scales.

Let us remind the reader of a framework we bring to deception judgments. In our view, people are harsher in evaluating others' lies than their own. They stereotype liars as conscience-stricken souls. When asked to judge deception, people consult this stereotype and assess its fit to the person at hand. In general, they are reluctant to label an assertion as deceptive when this judgment would imply that the person offering the assertion was a liar. The inaccurate stereotype and unwanted dispositional implication may help explain why receivers' judgments are so often inaccurate—more specifically, why so many deceptive messages are misclassified as truthful. Our double-standard hypothesis also provides a framework for interpreting the effects of various factors on deception judgments, effects which we now consider.

Deception medium. Deception can be judged over various media. Some may invite application of a stereotype for inferring deceit; others encourage reflection. The video medium, we suspect, should encourage use of a liar stereotype. Indeed, if forced to judge deceit from nothing more than a video image, observers have recourse to little other than their stereotypes. Access to verbal content gives judges the option of analyzing issues of veracity in a more thoughtful fashion. Thus, it is of interest to compare detection rates for lies that can only be seen versus those that can be heard.

Having sketched the relevance of our double-standard framework for interpreting deception attempts in different media, let us mention another theoretical perspective. According to Ekman and Friesen (1969), people should be most successful in their attempts at facial deceit and least successful in lying with the body because they are most motivated and able to control the face.

To assess the impact of deception medium, we identified 50 studies that experimentally manipulated this factor and extracted from these studies 177 pairwise comparisons of lie–truth discrimination accuracy in one medium versus another medium. Ninety eight of the comparisons were made on percentage correct lie–truth classifications and 79 on rating scales. Converting each comparison to a standardized mean difference, we conducted a fixed-effects meta-analysis. Results show that lie–truth discrimination accuracy is lower if judgments are made in a video rather than an audiovisual or audio medium (for comparison of video to audiovisual and audio lie–truth discrimination, weighted mean $ds = -.44$ and $-.37$, $Zs = -15.72$ and -9.51 in 58 and 34 experimental comparisons, respectively; each $p < .0001$). In fact, lie–truth discrimination from video presentations is inferior to discriminations made from written transcripts, weighted mean $d = -.28$, $Z = -4.16$, $p < .001$ in 10 experimental comparisons. The levels of lie–truth discrimination achieved from transcript, audiovisual, and audio presentations do not differ significantly from one another.

We tabulated analogous evidence of receivers' general tendency to perceive messages as truthful. Results show that messages are perceived as less truthful if judged from a video than an audiovisual or audio presentation, weighted mean $d = -.29$ and $-.34$, $Zs = -4.26$ and -5.79 in 14 and 15 experimental comparisons, respectively; each $p < .0001$. Messages conveyed in transcripts are judged as less truthful than audiovisual messages and as somewhat more truthful than those presented in video, weighted mean $ds = -.32$ and $.20$, $Z = -3.31$, $p < .01$ and $Z = 1.94$, $p = .06$, in 5 and 4 experimental comparisons, respectively. In perceived truthfulness, audio-based messages do not differ significantly from audiovisual or transcript messages; each $p > .10$.

To complement these within-study comparisons, we examined medium differences across all of the studies in the research literature. In 195 samples, we have data on percentage truth classifications to messages conveyed in one of three media: video only, audio only, or audiovisual. Relevant results appear in Table 2 and suggest that there is a truthfulness bias in judging messages that can be heard. Both audio-only and audiovisual presentations received more than 50% truth judgments. As the within-study comparisons indicated, video-only presentations are less often judged truthful. Medium effects on lie–truth discrimination appear in the bottom two thirds of the table. Corroborating the within-study evidence, these comparisons show that discrimination is poorer for video-only messages than for messages presented in an audio-only or audiovisual medium.

From our double-standard framework, we interpret these results as follows: that the usual stereotype of a liar is largely visual, hence is most strongly evoked by video images of people speaking. Those who can be viewed as tormented are judged to be lying, but apparent torment reflects many factors other than deceit.

Ekman and Friesen (1969) hypothesized that there are more deception cues in the body than the face. To examine this possibility, we divided the video-based deception attempts into ones that provided the receiver with cues from only the face ($k = 15$), only the body ($k = 9$), or the face plus the body ($k = 29$). Results provide only partial support for the Ekman and Friesen formulation. Consistent with that formulation, attempts at lie detection are unsuccessful when receivers see only the sender's face; however, detection efforts are similarly unsuccessful when receivers see only the liar's body (weighted mean accuracy $d = .01$, $-.15$, and $.12$ for face, body, and both, respectively).

Table 2. *Deception in Three Media: Within and Between Studies*

Within Studies			
Comparison	***k***	**Weighted Mean Accuracy *d* (95% CI)[a]**	
Video Versus Audio	34	−.371 (± .076)	Audio is more accurate
Video Versus Audiovisual	58	−.438 (± .053)	Audiovisual is more accurate
Audio Versus Audiovisual	47	−.056 (± .057)	
Between Studies			
	k	**Raw M[b] (95% CI)**	**Adjusted *M***
Total Truth Classifications			
Video	24	52.18% (± .54)	52.16%
Audio	24	58.78% (± .64)	63.38%
Audiovisual	147	56.32% (± .27)	56.20%
For the Difference		F_w (2, 4683) = 174.05 $p < .001$	
Correct Lie–Truth Classifications			
Video	37	50.52% (± .42)	50.35%
Audio	36	53.01% (± .43)	53.75%
Audiovisual	212	54.03% (± .22)	53.98%
For the Difference		F_w(2, 5348) = 118.38 $p < .001$	
Accuracy *d*			
Video	53	.077 (± .057)	.097
Audio	56	.419 (± .053)	.376
Audiovisual	278	.438 (± .022)	.448
For the Difference		$Q(2) = 132.17$ $p < .001$	$Q(2) = 140.04$ $p < .001$

[a]For within-study comparisons here and elsewhere, positive *d*s imply that lie/truth discrimination was higher in the condition listed first in the comparison; negative *d*s imply that it was higher in the condition listed second. In cases where the comparison is statistically significant (at $p < .05$), the condition that shows higher accuracy is noted in the table. [b]Percentages here and in later tables are precision weighted in the manner described by Bond, Wiitala, and Richard (2003).

Motivation. Deception studies are criticized when research participants have no incentive to be believed. Critics note that a lack of motivation may influence participants' believability. To address this issue, we divided the research literature into studies in which participants had little (or no) motivation to be believed and those in which they had higher motivation.

DePaulo and her colleagues (e.g., DePaulo et al., 1985) have hypothesized that senders are undermined by their efforts to get away with lying. In DePaulo's motivational impairment hypothesis, the truths and lies of highly motivated senders will be more easily discriminated than those of unmotivated senders unless receivers have access to nothing but a transcript of the sender's remarks.

For a controlled assessment of this hypothesis, we identified 20 studies that experimentally manipulated sender motivation, extracted from those studies 42 distinguishable motivation effects on lie–truth discrimination, and measured each effect as a standardized mean difference. Consistent with the motivational impairment hypothesis, experimental evidence shows that lies are easier to discriminate from truths if they are told by motivated rather than unmotivated senders (for impact of motivation, weighted mean $d = .171$, $Z = 7.10$, $p < .0001$).

The double-standard hypothesis has a different implication for understanding the impact of motivation on deception judgments. People who are afraid of being disbelieved may come to resemble the stereotypic liar. If so, they are likely to be judged deceptive. From this perspective, it should matter little whether or not a highly motivated speaker is lying. What matters is the speaker's fear of being disbelieved. High motivation would rarely make a person feel guilty or ashamed for lying; indeed, high stakes should make it easy to rationalize deceit.

For between-study evidence relevant to this perspective, Table 3. Consistent with the double-standard hypothesis, motivation to be believed reduces a speaker's apparent honesty. Perhaps motivation makes people resemble a visible stereotype of the liar. If so, motivational effects on credibility might be most apparent on video-based judgments. To assess this possibility, we examined the impact of motivation on lie- and truth-tellers' believability in video, audio, and audiovisual media. Between-study comparisons reveal that motivation significantly reduces senders' video and audiovisual appearance of truthfulness. For example, unmotivated and motivated senders are classified as truthful by 54.44% and 46.84% of receivers who see them in video-only presentations, $t'(95) = 7.17$, $p <$

Table 3. *Motivated and Unmotivated Deception Within and Between Studies*

Within Studies			
Comparison	***k***	**Weighted Mean Accuracy *d* (95% CI)**	
Motivated Versus Unmotivated	42	.171 (± .047)	Motivated is more accurate
Between Studies			
	k	**Raw *M* (95% CI)**	**Adjusted *M***
Total Truth Classifications			
No Motivation	130	57.24% (± .28)	57.19%
Motivation	85	53.43% (± .15)	55.66%
For the Difference		$t'(1,021) = 8.07$ $p < .001$	
Correct Lie–Truth Classifications			
No Motivation	177	53.36% (± .21)	53.43%
Motivation	125	53.85% (± .27)	53.27%
For the Difference		$t'(506) = 1.01$	
Accuracy *d*			
No Motivation	231	.462 (± .026)	.396
Motivation	170	.397 (± .028)	.371
For the Difference		$Q(1) = 10.80$ $p < .01$	$Q(1)=1.53$, n.s.

.001. However, motivation has no effect on how truthful a sender sounds, $t'(137) = 1.31$, n.s.

The bottom two thirds of Table 3 displays between-study evidence on sender motivation and lie–truth discrimination. Here it does not appear that motivation makes liars easier to detect.

Preparation. Sometimes the need to lie appears without warning, and people are unprepared for the deceptions they attempt. On other occasions, the need has been anticipated, and a line has been prepared. In principle, the opportunity to prepare might influence a liar's success.

To examine this possibility, we identified 15 studies that experimentally manipulated a sender's time to prepare lies. These studies reported 24 experimental effects of sender preparation on the accuracy of lie–truth judgments and 10 experimental effects on the sender's general tendency to appear truthful. A fixed-effects standardized meta-analysis shows that receivers achieve higher lie–truth detection accuracy when judging unplanned rather than planned messages (weighted mean $d = -.144$, $Z = 4.49$, $p < .01$), and that planned messages appear more truthful than unplanned messages (weighted mean $d = .133$, $Z = 2.35$, $p < .05$).

Relevant between-study evidence is displayed in Table 4. Although the results there for judgment accuracy are mixed, they suggest that it may be harder to discriminate deceptive from truthful messages when the messages are planned. Unlike the within-study evidence, between-study comparisons suggest that planned messages appear slightly less honest than spontaneous messages.

Baseline exposure to sender. The meta-analysis presented here focuses on judgments of deception among strangers. Even so, we included in the analysis 38 samples in which perceivers were exposed to a target before making judgments of that target. We also included 28 samples in which perceivers judged a given target eight or more times and 4 samples in which perceivers made a forced choice between a target's lie and that same target's truth. For purposes of the following analyses, all of these receivers were deemed to have received a baseline exposure to the target.

For a controlled analysis, we identified 21 experimental comparisons of the detection of a target's messages by judges who had (vs. judges who had not) been previously exposed to that target. All of these comparisons were made on percentage correct lie–truth judgments. Results indicate that baseline exposure improves lie–truth discrimination: Receivers achieve a mean of 55.91% accuracy when given a baseline exposure versus 52.26% accuracy in the absence of any exposure, $t'(364) = 6.37$, $p < .01$.

Between-study evidence on the impact of baseline exposure is displayed in Table 5. Results there suggest that baseline exposure may improve judgmental accuracy. At the same time, senders who are familiar to the receiver are likely to be given the benefit of the doubt, as results on the percentage of truth judgments indicates. Consistent with our double-standard framework, people are reluctant to imply that someone familiar to them is a liar.

Interaction. In many studies, people lie when alone or in the presence of a passive experimenter. In other studies, people are involved in social interactions when lying. Sometimes, the interaction partner is at-

Table 4. *Prepared and Unprepared Deceptions Within and Between Studies*

	Within Studies		
Comparison	***k***	**Weighted Mean Accuracy *d* (95% CI)**	
Prepared Versus Unprepared	24	−.144 (± .063)	Unprepared is more accurate
	Between Studies		
	k	**Raw *M* (95% CI)**	**Adjusted *M***
Total Truth Classifications			
No Preparation	118	56.33% (± .28)	57.18%
Preparation	99	55.49% (± .30)	55.15%
For the Difference		$t'(1130) = 1.96$ $p < .05$	
Correct Lie–Truth Classifications			
No Preparation	177	53.18% (± .21)	53.13%
Preparation	130	53.70% (± .26)	53.75%
For the Difference		$t'(506) = 1.13$	
Accuracy *d*			
No Preparation	217	439 (± .029)	.403
Preparation	184	.365 (± .028)	.361
For the Difference		$Q(1)=12.37$ $p < .001$	$Q(1)=4.10$ $p < .05$

Table 5. *Baseline Exposure to Sender Within and Between Studies*

	Within Studies		
Comparison	***k***	**Weighted Mean Accuracy *d* (95% CI)[a]**	
Exposure Versus No Exposure	21	.239 (± .091)	Exposure is more accurate
	Between Studies		
	k	**Raw *M* (95% CI)**	**Adjusted *M***
Total Truth Classifications			
No Exposure	187	56.11% (± .23)	55.31%
Exposure	31	58.37% (± .46)	61.92%
For the Difference		$t'(452) = 3.47$ $p < .01$	
Correct Lie–Truth Classifications			
No Exposure	250	53.35% (± .18)	53.06%
Exposure	61	54.22% (± .33)	54.55%
For the Difference		$t'(294) = 2.09$ $p < .05$	
Accuracy *d*			
No Exposure	331	.400 (± .022)	.356
Exposure	72	.443 (± .051)	.499
For the Difference		$Q(1) = 2.25$	$Q(1) = 32.12$ $p < .001$

[a]For within-study comparisons here and elsewhere, positive *ds* imply that lie/truth discrimination was higher in the condition listed first in the comparison; negative *ds* imply that it was higher in the condition listed second. In cases where the comparison is statistically significant (at $p < .05$), the condition that shows higher accuracy is noted in the table. [b]Percentages here and in later tables are precision weighted in the manner described by Bond, Wiitala, and Richard (2003).

tempting to judge the liar's veracity; on other occasions, a third party may be making this judgment. The latter occurs, for example, when the interaction partner is the experimenter and the third party is the receiver making judgments from a videotape. In principle, interaction might influence one's success at lying. Interaction might, for example, impose cognitive demands on the liar (Buller & Burgoon, 1996).

We found 11 studies that experimentally manipulated whether senders were interacting with the receiver or with a third party. Results indicate no significant difference in lie–truth discrimination by interaction partners (vs. third-party observers), weighted mean $ds = .286$ versus $.209$, $Z = 1.41$, n.s. We also tabulated evidence within 5 studies of receivers' general tendency to perceive senders as truthful. Results show that individuals are judged to be more truthful by their interaction partners than by third-party observers; for this comparison, weighted mean $d = .26$, $Z = 4.10$, $p < .0001$.

For between-study evidence on the impact of interaction, see Table 6. There it is again clear that receivers are inclined to judge their interaction partners as truthful. Overall patterns in the literature suggest that third-party observers are better than interaction partners at discriminating lies from truths. In our view, the reluctance to attribute deception to interaction partners results from an unwanted dispositional implication—of insinuating that the partner is a liar.

Receiver expertise. In most research, college students function as the judges of deception. Perhaps people who had more experience would be better at judging deceit. To assess this possibility, we identified studies of deception experts. These are individuals whose occupations expose them to lies. They include law enforcement personnel, judges, psychiatrists, job interviewers, and auditors—anyone whom deception researchers regard as experts.

In 19 studies, expert and nonexpert receivers judged the veracity of the same set of messages. From these studies, we extracted 20 independent expert–nonexpert comparisons and expressed each as a standardized mean difference. This cumulation yields no evidence that experts are superior to nonexperts in discriminating lies from truths; weighted mean $d = -.025$, 95% confidence interval = –.105 to.055. Indeed, the direction of the within-study difference favors higher nonexpert accuracy, though this difference is not statistically significant, $Z = -.61$, n.s. Within-study comparisons also reveal no statistically significant difference between experts and nonexperts in the tendency to perceive others as truthful; weighted mean percentage truth judgments = 54.09% and 55.74% for experts and nonexperts, respectively; $t'(246) = 1.41$.

For a broader assessment of experts' deception judgments, see Table 7. From the between-study evidence, it would appear that experts are more skeptical than nonexperts, being less inclined to believe that people are truthful. Having been targets of deceit in their professional roles, experts may have surmounted the usual reluctance to imply that people are liars. If raw between-study comparisons suggest that experts may be better than nonexperts at discriminating lies from truths, it is clear that experts are not good lie detectors. On the average, they achieve less than 55% lie–truth discrimination accuracy. In any case, experts' apparent superiority in lie–truth discrimination disappears when means are statistically adjusted.

Publication status. Lie detection results might influence the likelihood of a research project being published. To assess this possibility, we did a few other analyses. These reveal no statistically significant differences between published and unpublished studies in lie–truth discrimination performances. For example,

Table 6. *Sender Interaction Within and Between Studies*

Within Studies

Comparison	k	Weighted Mean Accuracy d (95% CI)[a]
Interaction With Receiver Versus Third Party	10	.081 (± .094)

Between Studies

	k	Raw M (95% CI)	Adjusted M
Total Truth Classifications			
No Interaction	66	54.51% (± .34)	57.58%
Interaction With Receiver	13	65.32% (± 2.05)	61.60%
Interaction With Third Party	128	55.51% (± .28)	56.27%
For the Differences	$F_w(2, 1403) = 58.15$ $p < .001$		
Correct Lie–Truth Classifications			
No Interaction	85	52.56% (± .27)	52.60%
Interaction With Receiver	18	52.27% (± 1.68)	52.75%
Interaction With Third Party	189	54.06% (± .20)	53.97%
For the Differences	$F_w(2, 2051) = 37.67$ $p < .001$		
Accuracy d			
No Interaction	127	.375 (± .036)	.302
Interaction With Receiver	33	.234 (± .076)	.316
Interaction With Third Party	224	.416 (± .027)	.471
For the Differences		$Q(2) = 20.24$ $p < .01$	$Q(2)=57.14$ $p < .001$

[a]For within-study comparisons here and elsewhere, positive ds imply that lie/truth discrimination was higher in the condition listed first in the comparison; negative ds imply that it was higher in the condition listed second. In cases where the comparison is statistically significant (at $p < .05$), the condition that shows higher accuracy is noted in the table. [b]Percentages here and in later tables are precision weighted in the manner described by Bond, Wiitala, and Richard (2003).

Table 7. *Receiver Expertise Within and Between Studies*

Within Studies			
Comparison	***k***	**Weighted Mean Accuracy *d* (95% CI)[a]**	
Expert Versus Nonexpert	20	–.025 (± .080)	
Between Studies			
	k	**Raw *M* (95% CI)**	**Adjusted *M***
Total Truth Classifications			
Nonexpert	177	55.69% (± .20)	55.84%
Expert	30	52.28% (± .58)	52.02%
For the Difference		$t'(361) = 4.95$ $p < .001$	
Correct Lie–Truth Classifications			
Nonexpert	250	53.31% (± .17)	53.29%
Expert	42	54.51% (± .47)	53.81%
For the Difference		$t'(556) = 2.37$ $p < .05$	
Accuracy *d*			
Nonexpert	338	.380 (± .022)	.387
Expert	46	.488 (± .064)	.388
For the Difference		$Q(1) = 9.77$ $p < .01$	$Q(1)=.01$

Note: Expert receivers have a background researchers deem relevant to detecting deception. They include police officers, detectives, judges, interrogators, criminals, customs officials, mental health professionals, polygraph examiners, job interviewers, federal agents, and auditors. Percentages are precision weighted in the manner described by Bond, Wiitala, and Richard (2003).
[a]For within-study comparisons here and elsewhere, positive *d*s imply that lie/truth discrimination was higher in the condition listed first in the comparison; negative *d*s imply that it was higher in the condition listed second. In cases where the comparison is statistically significant (at $p <$ .05), the condition that shows higher accuracy is noted in the table.

the weighted mean percentage correct lie–truth classifications is 53.19% in published studies and 53.75% in unpublished studies, $t'(872) = 1.49$, n.s. Truthfulness biases were, however, stronger in unpublished research; weighted mean percentage truth classifications = 56.75% versus 54.27% in unpublished versus published research, $t'(498) = 4.75$, $p < .001$.

Discussion

Having captivated human imagination for millenia, deception was destined to attract psychological investigators. Our goal has been to synthesize their research—more specifically, to quantify people's ability to detect deceit from behavior. Here we summarize the findings of our meta-analysis, discuss the literature in light of a double-standard framework, and note limitations in the existing evidence.

Meta-Analytic Findings

How successful are people at duping others? How often do people detect others' deception attempts? To address these questions, psychologists arrange for people to make truthful and deceptive statements and for others to classify these statements as truths or lies. Across hundreds of experiments, typical rates of lie–truth discrimination are slightly above 50%. For the grand mean, 54% is a reasonable estimate.

Having noted that the average person discriminates lies from truths at a level slightly better than he or she could achieve by flipping a coin, let us also note this ability corresponds to a nontrivial standardized effect size. In producing a mean difference of approximately .40 standard deviations in judgments of lies versus truths, typical detection abilities are larger than 60% of the research phenomena studied by social psychologists (Richard et al., 2003).

Our finding of a 54% lie–truth discrimination rate represents an average of correct judgments to deceptive messages and truthful messages. It is clear that truthful messages are more often judged correctly than deceptive messages; hence, the percentage of correct judgments to messages encountered in any real-world setting may depend on the base rate of deception there. In a setting where virtually no lies were told, the research literature would suggest a detection rate of roughly 60%, whereas in a situation where virtually every statement was a lie, a detection rate of, say, 48% might be expected (cf. Levine, Park, & McCornack, 1999). These estimates assume that there is no cross-situational correlation between observers' tendency to infer deception in a setting and the actual rate of lying there. More likely, deception base rates enter into a tactical calculus. As observers have intuitions about the frequency of deception in different situations, liars have intuitions, too. If the latter can choose where to attempt their deceptions, they should opt for settings in which targets are most trusting.

Like earlier reviewers, we find that people are more inclined to judge deceptive messages as truthful than truthful messages as deceptive. No doubt, receivers contribute to this truth bias, but senders' contributions should also be acknowledged. When people try to appear truthful, their efforts are rewarded, the accumulated literature shows. The relative impact of senders and receivers on the truth bias remains to be determined. In the meantime, the present contribution is to document the magnitude of this effect. Across 206 studies, people render a mean of some 56% truth judgments. However, this figure may understate the presumption of truth telling in real life. If in their daily interactions people accept without reflection much of what they hear, in the laboratory they are forced to make veracity judgments. Thus, researchers circumvent some of the usual impediments to inferring deceit—social norms that discourage skepticism, liars' tactics for preempting suspicion, and a cognitive inertia that would be disrupted by critical inquiry (Levine et al., 1999).

We see a pattern in this research literature. In their reading of the literature, scholars find an unwanted implication—that people can barely discriminate lies from truths. Heirs to the moralistic tradition, scholars resist this implication by identifying a feature of researchers' methods that could in principle explain low lie–truth discrimination rates. They label the feature an *artifact*, correct the error, run a study, and announce that their findings are uniquely valid. Sometimes, the methodological correction yields a higher than average detection rate, and sometimes it does not. Never, however, has this quest for accuracy yielded levels of lie detection that would be of much practical use. Occasionally, a researcher finds a detection rate of 70% (or so) and proclaims a momentous discovery. However, those rates occur on tests that include only a small number of messages and are attained by only a subset of the receivers (or on a subset of the tests) studied. From a meta-analytic perspective, random variation is the most plausible explanation for the occasionally high detection rate, as the funnel pattern in Figure 2 suggests.

Rather than marveling at the outliers in this literature, we are more impressed by the regularity of the results obtained. Despite decades of research effort to maximize the accuracy of deception judgments, detection rates rarely budge. Professionals' judgments, interactants' judgments, judgments of high-stakes lies, judgments of unsanctioned lies, judgments made by long-term acquaintances—all reveal detection rates within a few points of 50%. We wonder if it is premature to abort the quest for 90% lie detection and accept the conclusion implied by the first 384 research samples—that to people who must judge deception in real time with no special aids, many lies are undetectable.

Although rates of lie detection vary within a narrow range, the variation is not random. Some factors facilitate lie–truth discrimination, and others impede it, our meta-analytic results confirm. The medium in which deception is attempted affects its likelihood of detection—lies being more detectable when they can be heard. By contrast, facial behaviors provide no indication of a speaker's veracity, corroborating the theory that the face is well controlled (Ekman & Friesen, 1969). Ekman and Friesen also suggested that bodily behaviors go uncontrolled, and hence should be indicative of deceit. Unfortunately, the latter hypothesis has so rarely been tested that its validity remains unknown.

A more recent perspective (Buller & Burgoon, 1996) emphasized the role of social interaction in deception judgments. The accumulated research suggests that lies told in the midst of social interaction are spotted by onlookers, yet they are fooling the liar's interaction partner. However, controlled experiments show no difference in lie detection by interaction partners as opposed to onlookers. As common sense might have predicted, judges achieve better lie–truth discrimination if they have a baseline exposure to the sender and if the sender is unprepared. The accumulated evidence suggests that people who are motivated to be believed look deceptive whether or not they are lying. Expert judges may be slightly more skeptical than novices. Relative to novices, experts may (or may not) be better at lie–truth discrimination; in any case, they make many mistakes.

The Double Standard

Having reviewed the research literature on deception judgments and cataloged some factors that influence detection accuracy, let us note the relevance of our favored framework for understanding this subject—our assumption that people judge others' deceptions more harshly than their own.

We do not regard this meta-analysis as a test of the notion of a double standard. In our view, no test for so obvious an idea is needed—though relevant evidence can be found in primary research (Gordon & Miller, 2000). Instead, we begin with the premise that people construe others' lies more critically than their own and explore the implications of this premise for understanding research findings.

Indignant at the prospect of being duped, people project onto the deceptive a host of morally fueled emotions—anxiety, shame, and guilt. Drawing on this stereotype to assess others' veracity, people find that the stereotype seldom fits. In underestimating the liar's capacity for self-rationalization, judges' moralistic stereotype has the unintended effect of enabling successful deceit. Because deceptive torment resides primarily in the judge's imagination, many lies are mistaken for truths. When torment is perceived, it is often not a consequence of deception but of a speaker's motivation to be believed. High stakes rarely make people feel guilty about lying; more often, they allow deceit to be easily rationalized.

When motivation has an impact, it is on the speaker's fear of being disbelieved, and it matters little whether or not the highly motivated are lying. The impact of motivation is most evident when judges can see the speaker's resemblance to a visual stereotype of the liar.

People are critical of lies, unless the lies are their own. To maintain an exception for themselves, judges may sometimes need to excuse lying by others. As the research literature shows, people avoid attributing deception to others with whom they are familiar—whether from a live interaction or a long-term relationship (Anderson, Ansfield, & DePaulo, 1999). Judges may also be loath to perceive as liars people who resemble the judge. Perhaps the truth bias we observe in this literature represents an extension of the self-bias to others who are reminiscent of the self. In this view, the bias reflects the similarity of the deceivers in this research literature to their judges—often, the two are students at the same University. Maybe there would be less bias in judgments made of dissimilar others. As we have noted, deception researchers find that expert judges are willing to imply that others are liars. What we have not noted is a procedural detail—that these experts are rarely sitting in judgment of their peers; instead, they are judging members of other groups. Self-biases do not extend to outsiders.

The judges in this research literature are given the goal of achieving 100% accuracy, and their failure to attain this objective has been widely lamented. The senders in this research literature are also given a goal: to convey an impression of honesty 100% of the time. Results show that research participants disbelieve nearly 50% of senders' deception attempts and nearly 40% of their attempts at truth telling. Although in the rough actuarial aggregate of deception research liars fail as often as detectors, deception failures have rarely been discussed. Let us comment on these failures from a double-standard perspective.

Liars who are often judged deceptive should come to learn that their stratagems have been penetrated. Thus, it may seem paradoxical that the average person lies several times a day (DePaulo, Kashy, et al., 1996). Evidently, most lies are little, and the consequences of detection benign. In the interest of interacting smoothly, the liar and judge conspire to preserve a fiction (DePaulo et al., 2003).

A few lies involve high stakes: large benefits to the liar and large costs to the dupe. Moralists focus on these big bad lies. The research literature has explored judgments made at the time deception is attempted, judgments that could preempt the payoffs liars pursue. However, research reveals that many people avoid being caught in the act of lying; hence, it is important to explore the likely course of subsequent events.

High-stakes deceptions are motivated by noncorrespondent outcomes, one person seeking advantage at another's expense. There are costs of being duped, and these should impose limits on the dupe's naiveté. Some lies are discovered well after they have been told (Park, Levine, McCornack, Morrison, & Ferrara, 2002). Then the dupes become indignant. They retaliate by shunning their exploiter and publicizing the liar's duplicity. As a consequence, people who are most successful in the short-term perpetration of lies have trouble maintaining relationships. Moralists have opined that skilled liars are worse relationship partners than highly honest folk. Let us suggest that skilled liars may also be worse partners than people whose lies are transparent (Andrews, 2002). Inept liars pose no threat to their partners insofar as their deception attempts fail before any damage is done. This line of reasoning suggests that skill at high-stakes deception may be an interpersonal liability and that so-called deception failures are in the long run adaptive.

Maybe the craftiest can benefit from lying. Cognizant of the dispositional nature of moral attributions (Reeder, 1993), they cultivate reputations for honesty by telling the truth on trivial matters and noting advantages that fibbing might have conferred. Then, when deceit promises the largest reward, others will have been lulled into an unwarranted trust (Sternglanz, 2003). Having laid the tactical groundwork, liars must nonetheless recognize that deceptions may ultimately be exposed. In the moment of lying, the shrewdest affect a distancing from their falsehoods so that they can later disavow the lies. For deception to show long-term profitability, reputational damage must be contained.

Limitations in the Evidence

Commentators have criticized research on deception judgments, pointing to ways in which the lies studied in the research literature differ from the lies of most interest to the critic. Those who are interested in high-stakes lies (Ekman, 2001) note that many experimental deceptions are trivial. Those who are interested in deceptive interactions (Buller & Burgoon, 1996) denounce experimentally constrained lies. Legal scholars (e.g., Fisher, 1997) note aspects of the forensic world that are not reproduced in research contexts.

Deception researchers have tried to accommodate critics' reservations. They have studied murderers' lies and lies that could harm children (Lusby, 1999; Vrij & Mann, 2001), lies to lovers and deceit during criminal interrogations (Anderson et al., 1999; Davis, Markus, Walters, Vorus, & Conners, 2005). Researchers have studied naturalistic deceptive interactions and jurors' credibility judgments. In light of these efforts, we find no merit in blanket dismissals of this research literature as trivial, asocial, and irrelevant.

We ourselves have reservations about the literature on deception judgments, concerns that have not (we think) been addressed. To illuminate lie detection from language and behavior, psychologists have excluded from their research other potential cues to deception. They have restricted the time span over which issues of deception can be pondered, blinded judges to the motivational contingencies surrounding deceit, and neutralized naturally occurring correlates of the propensity to lie.

In experiments, judges encounter a message and must judge the veracity of that message on the spot, with no time to gather additional information. Outside the laboratory, additional information is important. When asked to describe their discovery of a lie, people rarely state that the discovery was prompted by behaviors displayed at the time of the attempted deception. Rather, they say that lie detection took days, weeks, or even months, and involved physical evidence or third parties (Park et al., 2002). Surely, motivational information conditions real-world deception judgments—when, for instance, jurors discount expert testimony after learning that the expert received a fee (Hilton, Fein, & Miller, 1993). In venues of frequent deception, people may base their veracity judgments more strongly on perceived incentives than any behavioral information. People differ widely in the propensity to lie (Kashy & DePaulo, 1996), and this individual difference may be discernable (Bond, Berry, & Omar, 1994). Researchers bypass naturally occurring correlates of deceptiveness by compelling lies from every experimental participant—even those who are loath to lie. Future studies will be needed to explore the impact on lie detection of these and other forms of extra-behavioral information. Perhaps the 90% lie detection barrier will someday be broken.

In the meantime, we have accumulated knowledge about judgments of deception from speech content and behavior. Yes, people often fail in their efforts to divine deception, and this raises questions about the American legal system, in which jurors are responsible for detecting lies. It is important also to note that research participants often fail when trying to dupe others. Perhaps it would be unsurprising if liars and would-be detectors had arrived at an equilibrium. If liars were much better, truth telling would be less common; if detectors were much better, few lies would be attempted.

References

Anderson, D. E., Ansfield, M. E., & DePaulo, B. M. (1999). Love's best habit: Deception in the context of relationships. In P. Philippot & R. S. Feldman (Eds.), *Social context of nonverbal behavior* (pp. 372–409). New York: Cambridge University Press.

Anderson, N. H. (1968). Likeableness ratings of 555 personality-trait words. *Journal of Personality and Social Psychology, 9,* 272–279.

Andrews, P. W. (2002). The influence of postreliance detection on the deceptive efficacy of dishonest signals of intent: Understanding facial clues to deceit as the outcome of signaling tradeoffs. *Evolution and Human Behavior, 23,* 103–121.

Bok, S. (1989). *Lying: Moral choice in public and private life.* New York: Vintage.

Bond, C. F., Jr., Berry, D. S., & Omar, A. (1994). The kernel of truth in judgments of deceptiveness. *Basic and Applied Social Psychology, 15,* 523–534.

Bond, C. F., Jr., Wiitala, W., & Richard, F. D. (2003). Meta-analysis of raw mean differences. *Psychological Methods, 8,* 406–418.

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

Davis, M., Markus, K. A., Walters, S. B., Vorus, N., & Connors, B. (2005). Behavioral cues to deception vs. topic incriminating potential in criminal confessions. *Law and Human Behavior, 29,* 683–704.

DePaulo, B. M. (1992). Nonverbal behavior and self-presentation. *Psychological Bulletin, 111,* 203–243.

DePaulo, B. M., Ansfield, M. E., & Bell, K. L. (1996). Theories about deception and paradigms for studying it: A critical appraisal of Buller and Burgoon's interpersonal deception theory and research. *Communication Theory, 3,* 297–310.

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., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin, 129,* 74–118.

DePaulo, B. M., & Pfeifer, R. L. (1986). On-the-job experience and skill at detecting deception. *Journal of Applied Social Psychology, 16,* 249–267.

DePaulo, B. M., Stone, J. L., & Lassiter, G. D. (1985). Deceiving and detecting deceit. In B.R. Schenkler (Ed.), *The self and social life* (pp. 323–370). New York: McGraw-Hill.

DePaulo, B. M., Zuckerman, M., & Rosenthal, R. (1980). Humans as lie detectors. *Journal of Communication, 30,* 129–139.

Ekman, P. (2001). *Telling lies: Clues to deceit in the marketplace, politics, and marriage* (3rd ed.). New York: Norton.

Ekman, P., & Friesen, W. V. (1969). Nonverbal leakage and clues to deception. *Psychiatry, 32,* 88–105.

Fisher, G. (1997). The rise of the jury as lie detector. *Yale Law Journal, 103*, 575–713.

Fleiss, J. L. (1994). Measures of effect size for categorical data. In H. Cooper and L. Hedges (Eds.), *The handbook of research synthesis* (pp. 245–260). New York: Russell Sage Foundation.

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

Gordon, A. K., & Miller, A. G. (2000). Perspective differences in the construal of lies: Is deception in the eye of the beholder? *Personality and Social Psychology Bulletin, 26,* 46–55.

Hedges, L. V., & Vevea, J. L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods, 3,* 486–504.

Hilton, J. L., Fein, S., & Miller, D. T. (1993). Suspicion and dispositional inference. *Personality and Social Psychology Bulletin, 19,* 501–512.

Hubbell, A. P., Mitchell, M. M., & Gee, J. C. (2001). The relative effects of timing of suspicion and outcome involvement on biased message processing. *Communication Monographs, 68,* 115–132.

Judicial Committee on Modern Jury Instructions for the Eighth Circuit. (2002). *Manual of modern criminal jury instructions.* Retrieved May 15, 2006, from http://www.are.uscourts.gov/jury/crjury2002.pdf

Kalbfleisch, P. J. (1990). Listening for deception: The effects of medium on accuracy of detection. In R. N. Bostrom (Ed.), *Listening behavior: Measurement and application* (pp. 155–176). New York: Guilford.

Kashy, D. A., & DePaulo, B. M. (1996). Who lies? *Journal of Personality and Social Psychology, 70,* 1037–1051.

Kraut, R. (1980). Humans as lie detectors: Some second thoughts. *Journal of Communication, 30,* 209–216.

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.

Light, R. J., Singer, J. D., & Willett, J. B. (1994). The visual presentation and interpretation of meta-analyses. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 439–453). New York: Russell Sage Foundation.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis.* Thousand Oaks, CA: Sage

Lusby, David J. (1999). *Effects of motivation to lie and sanctioned deception of the accuracy of observers' veracity judgments*. Unpublished master's thesis, SUNY-Buffalo, New York.

Mattson, M., Allen, M., Ryan, D. J., & Miller, V. (2000). Considering organizations as a unique interpersonal context for deception detection: A meta-analytic review. *Communication Research Reports, 17,* 148–160.

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

O'Sullivan, M. (2003). The fundamental attribution error in detecting deception: The boy-who-cried-wolf effect. *Personality and Social Psychology Bulletin, 29,* 1316–1327.

Park, H. S., Levine, T. R., McCornack, S. A., Morrison, K., & Ferrara, S. (2002). How people really detect lies. *Communication Monographs, 69*, 144–157.

Reeder, G. D. (1993). Trait–behavior relations and dispositional inference. *Personality and Social Psychology Bulletin, 19*, 586–593.

Richard, F. D., Bond, C. F., Jr., & Stokes-Zoota, J. J. (2003). One hundred years of social psychology quantitatively described. *Review of General Psychology, 7*, 331–363.

Saxe, L. (1991). Lying: Thoughts of an applied social psychologist. *American Psychologist, 46,* 409–415.

Sternglanz, R. W. (2003). *Exoneration of serious wrongdoing via confession to a lesser defense*. Unpublished doctoral dissertation, University of Virginia, Charlottesville.

Swets, J. A. (1996). *Signal detection theory and ROC analysis in psychology and diagnostics*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Trovillo, P. V. (1939). A history of lie detection. *Journal of Criminal Law and Criminology, 29*, 848–881.

Vrij, A. (2000). *Detecting lies and deceit: The psychology of lying and the implications for professional practice*. New York: Wiley.

Vrij, A., & Mann, S. (2001a). Telling and detecting lies in a high-stake situation: The case of a convicted murderer. *Applied Cognitive Psychology, 15*, 187–203.

Walter, S. D. (2002). Properties of the summary receiver operating characteristic (SROC) curve for diagnostic test data. *Statistics in Medicine, 21*, 1237–1256.

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. 1–60). New York: Academic.