Personality and Social Psychology Review
1997, Vol. 1, No. 4, 346–357

The Accuracy–Confidence Correlation in the Detection of Deception

Bella M. DePaulo
Department of Psychology
University of Virginia

Kelly Charlton, Harris Cooper, James J. Lindsay, and Laura Muhlenbruck
Department of Psychology
University of Missouri, Columbia

A meta-analysis was conducted of research on the relation between judges' accuracy at detecting deception and their confidence in their judgments. A total of 18 independent samples revealed an average weighted accuracy–confidence correlation of .04, a relation not significantly different from zero. However, confidence was positively correlated with judges' tendency to perceive messages as truthful, regardless of the actual truthfulness of the messages. Judges were also more confident when they really were rating truths compared to when they were rating lies. Also, men were more confident than women, and judges who had a closer relationship to the message sender felt more confident in their judgments of truths and lies. Methodological and theoretical explanations for these findings are discussed.

Dozens of studies of the communication of deception provide compelling evidence that people are not very skilled at distinguishing when others are lying from when they are telling the truth. In experimental studies of detecting deception, accuracy is typically only slightly better than chance (DePaulo, Stone, & Lassiter, 1985; Zuckerman, DePaulo, & Rosenthal, 1981). Even in studies of people who have worked for years at jobs that involve the detection of deception, such as customs inspectors, it is rare to find impressive levels of accuracy (DePaulo & Pfeifer, 1986; Ekman & O'Sullivan, 1991; Kraut & Poe, 1980).

If people are poor judges of when deceit occurs, then they ought not trust their own sense of whether or not they are being deceived. However, there may be times when people feel very sure of their judgments, and other times when they know they are just guessing. If these feelings of confidence are linked to accuracy, then people can use confidence as a guide. That is, they can trust their judgments of truth and deceit to be correct when they feel sure that they are correct, and they can know not to trust their judgments when they feel unsure.

This research was supported in part by National Science Foundation Grant SBR–931 9556.

Requests for reprints should be sent to Bella DePaulo, Department of Psychology, Gilmer Hall, University of Virginia, Charlottesville, Virginia 22903. E-mail: bmd@virginia.edu.

The Accuracy–Confidence Correlation

Is confidence a reliable guide to accuracy in other domains of human judgment and decision making other than the detection of deception? In cognitive psychology, decades of research point to an "overconfidence" effect: People feel more confident than is warranted by measures of their actual competence (Fischhoff, Slovic, & Lichtenstein, 1977; Koriat, Lichtenstein, & Fischhoff, 1980). This overconfidence effect is especially characteristic of tasks that are moderately or very difficult. On easy tasks, people sometimes miss the mark in the opposite direction: They are less confident than is warranted by the facts (e.g., Lichtenstein & Fischhoff, 1977).

Griffin and Tversky (1992) argued that overconfidence (and underconfidence) effects can be explained by perceivers' overreliance on the strength of the available evidence and their underreliance on its weight or credibility. For example, in admissions decisions, people will inappropriately have more confidence in a candidate who has a very strong letter from a writer who has known the candidate for only a short time than they will in another candidate with a moderately strong letter from a writer who has known the candidate for many years.

We believe that perceivers attempting to detect deception will also show inappropriately high levels of confidence. Deception detection is a very difficult task; accuracy is typically only a few percentage points better than chance (e.g., DePaulo et al., 1985). In addition, it

is a task at which perceivers may weigh too heavily cues that have little or no predictive utility.

Previous research (reviewed in DePaulo et al., 1985, and Zuckerman, DePaulo, et al., 1981) has documented cues that really do occur more or less frequently when people are lying compared to when they are telling the truth (actual cues to deception), as well as cues that people regard as indicative of deception (perceived cues to deception). An important conclusion of that research is that actual and perceived cues to deception are not always the same (Fiedler & Walka, 1993; Zuckerman, Koestner, & Driver, 1981). Although there is some correspondence between actual and perceived cues, people sometimes use cues that have no reliable relation to actual deception, and they sometimes ignore cues that are valid indices of deceit. Another important conclusion is that even those cues that really are linked to deceit are linked only probabilistically. There is no behavior that always occurs when people are lying and never occurs when they are telling the truth.

When judges trying to detect deception see a cue or set of cues that they believe to be associated with lying, they may become overconfident. They may not realize that the cue is not in fact indicative of deceit or is not as reliably associated with deceit as they believe.

In this article, we review the evidence for overconfidence effects in the detection of deception, but our primary focus is on the correlation between confidence and accuracy. Perceivers' flawed theories of cues to deception are also pertinent to the accuracy–confidence relation. When perceivers use invalid (misleading) or irrelevant cues, their confidence may increase while their accuracy declines or remains unaffected. When perceivers overlook valid cues, their confidence may be unaffected but their accuracy may decline. When perceivers notice cues that really are indicative of deceit, both confidence and accuracy will increase. Because the cues that perceivers use include valid, invalid, and irrelevant ones, we think that there will be no significant overall relation between confidence and accuracy.

That prediction is consistent with the literature on accuracy of interpersonal judgments in domains other than deception detection. For example, in Ickes's (1993) review of the literature on empathic accuracy, which is the "ability to infer the specific content of another person's thoughts and feelings" (p. 588), he concluded that people have little or no metaknowledge about their own empathic skill. Perceivers' self-reports of their empathic accuracy are uncorrelated, or even slightly negatively related, to their actual accuracy. In a study in which perceivers' confidence in each of their empathic judgments was directly assessed and then correlated with the accuracy of those judgments, the result was again a nonsignificant (and slightly negative) relation (Thomas, Fletcher, & Lange, 1997).

Ickes (1993) speculated that an important impediment to the development of metaknowledge about empathic ability is the paucity of feedback. Perceivers rarely ask if their empathic inferences are correct, and targets rarely volunteer the relevant information. In fact, targets may even act in ways that are misleading, as, for example, when they nod to be polite even though they actually disagree. In the domain of deception, these difficulties are magnified. In everyday social interactions, perceivers rarely state their suspicions, and even when they do, targets rarely confirm them. There are, of course, contexts such as security checkpoints in airports in which perceivers (customs inspectors) have the explicit goal of attempting to detect deception (Kraut & Poe, 1980). However, even in those contexts, feedback is not systematic (DePaulo & Pfeifer, 1986). Inspectors can fail to search travelers who were in fact smuggling contraband, and they can even search a traveler and fail to find hidden contraband. The behavior of the travelers in both instances might contribute to the inspectors' notions of how truthful travelers act (and to their feelings of confidence when they see the behavior again), when instead the behavior should be informing impressions of deceptiveness.

In the literature on eyewitness testimony, research and theory on the relation between accuracy and confidence is already quite advanced. That literature reveals a generally negligible relation, except under certain conditions (Deffenbacher, 1980; Kassin, 1985; Wells & Murray, 1984). The accuracy–confidence link can be bolstered by giving the witness favorable information-processing conditions, such as having more time to see the perpetrator's face. With more time, the memory trace might be strengthened, and both accuracy and confidence might improve (Bothwell, Deffenbacher, & Brigham, 1987).

We do not expect the same qualifications to occur in the domain of deception detection. In that realm, more propitious information-processing conditions are no guarantee of greater accuracy because detection is not a matter of recognition but of theory and inference. If people's theories about the kinds of cues that indicate deception are erroneous, then it will be of little use to them to have more time to collect data relevant to those theories or to have better memories of the actual behaviors of the liars and truthtellers.

More generally, accuracy and confidence may show little relation because they follow from processes that are not always the same. For example, we know that experience at jobs that involve detection of deception adds little to accuracy (e.g., DePaulo & Pfeifer, 1986; Ekman & O'Sullivan, 1991; Kraut & Poe, 1980). However, if people in those jobs believe that their experience has helped them or should help them, then their confidence may be augmented by their job experience.

In sum, then, our prediction that accuracy and confidence in detecting deception would not be significantly related is based on evidence suggesting that (a) the accuracy–confidence link is sometimes negligible

in other domains, (b) people appear to hold erroneous theories about cues to deception, and (c) judgments of deception and ratings of confidence may follow from different processes.

Other Correlates of Confidence

Although we predicted that confidence would not be linked to accuracy, there is reason to believe confidence would be systematically related to other theoretically important variables.

Truth bias. The literature on the detection of deception has shown that judges (perceivers) tend to believe that others are telling the truth (DePaulo et al., 1985; Zuckerman, DePaulo, et al., 1981). Even when they are watching senders who are lying and telling the truth equally often, they report that more of the messages are truths than lies. In this review, we addressed the question of whether the truth bias is related to judges' confidence.

McCornack and his colleagues (Levine & McCornack, 1992; McCornack & Parks, 1986) developed a model of deception detection in personal relationships that predicts a positive correlation between confidence and truth bias. They believe that as relationships develop and become closer, relationship partners will become more confident about their judgments of each other's truths and lies. Then, as confidence increases, the partners will be more likely to see more of each other's communications as truths. Because McCornack and his colleagues have already found some support for their model, we predicted that the results of our review would also show a positive relation between confidence and truth bias. We also tested the first step in their model, which predicts that relationship closeness will be positively correlated with confidence.

Actual truthfulness of the messages. DePaulo (1994) suggested that although people are generally not very accurate at distinguishing truths from lies when they are asked directly whether particular messages are truths or lies, they may be able to distinguish the two kinds of messages in more indirect ways. For example, they may perceive deceptive messages as more ambivalent than truthful ones, or they may feel more confident in their judgments when they are observing truths than when they are observing lies. In this review, we tested the prediction that judges would feel more confident in their judgments of truths than of lies. That is, apart from whether people judge a message to be truthful or deceptive, and also apart from whether that judgment is correct, people may be substantially more confident when they are judging actual truths than when they are judging actual lies.

Methods

Overview

This review synthesizes research on the relation between accuracy and confidence in the detection of deception. First, we searched the literature for all studies that correlated a measure of accuracy at detecting deception with a measure of confidence in those judgments. Then, we used meta-analytic procedures to combine the results. Finally, we statistically examined potential moderators of the accuracy–confidence relation (see Cooper & Hedges, 1994).

Search Outcomes

The literature search uncovered 36 reports of empirical studies that included a measure of self-reported confidence in judgments of deceptiveness. Of those, 28 contained sufficient information for the calculation of effect sizes. Most relevant to the present investigation were 16 sources that reported correlations between the confidence measure and a measure of accuracy at detecting deception, but reports of other correlations with confidence were also of interest.

Studies were obtained from the first author's files; solicitation of reports from authors doing research on deception (68 letters were sent out, and some of these were passed on to other colleagues); from examining the reference lists in sources on the communication of deception; and from PsycLIT (American Psychological Association, 1974–present) searches using key words such as deception, lying, and lies.

Methods of Meta-Analysis

Metrics for expressing effect sizes. For each sample we recorded or calculated correlation coefficients between confidence and a second variable. The majority of the effect sizes were the result of relating two continuous variables. For the other effects that were derived from comparisons between two groups, a *d*-index was first calculated and then converted to a correlation (Cohen, 1988).

Calculation of average effect sizes. Both weighted and unweighted procedures were used to calculate average correlations across samples. First, correlations were trans-

formed to z scores so as to normalize their distributions. Then, in the unweighted procedure, each independent z score was given equal weight in calculating the average effect. In the weighted procedure, each independent z score was multiplied by its corresponding sample size minus three, and then this sum was divided by the sum of the weights. In this procedure, effect sizes based on larger samples are given greater weight. Finally, average z-scores were transformed back into average correlation coefficients.

One problem that arises in calculating average effect sizes is deciding what constitutes an independent estimate of effect. In this meta-analysis we used a shifting unit approach (Cooper, 1989). Specifically, each effect size is first recorded as if it were an independent event. So, for example, if a single sample description permitted separate calculations for the relation between accuracy and both a prejudgment and a postjudgment confidence measure, the two correlations would be coded separately. However, for the overall estimate of the relation between confidence and accuracy, these two effect sizes would be averaged prior to entry into the analysis. In the weighted procedure, this single (averaged within sample) correlation would be weighted by its sample size. If a subsequent analysis examined prejudgments versus postjudgments of confidence in relation to accuracy, the sample would contribute one effect estimate to each of the two calculations.

The shifting unit approach retains as much data as possible while holding to a minimum violations of the assumption that data points are independent. In this analysis, we identified 76 correlations drawn from 31 independent samples (described in the 28 research reports).

Whether average correlations were reliably different from 0 was tested by calculating 95% confidence intervals for weighted estimates. If the confidence interval did not contain $r = .00$, then the null hypothesis was rejected.

Tests for moderators of effects involving confidence. Possible moderators of relations involving confidence were tested using homogeneity analysis (Cooper & Hedges, 1994; Hedges & Olkin, 1985). Homogeneity analysis compares the amount of variance exhibited by a set of effect sizes with the amount expected if only sampling error is operating. The analysis can be carried out within groups of individual effect sizes (e.g., the entire set of correlations) or between group averages (e.g., the average correlation for prejudgment vs. postjudgment confidence ratings).

Hedges and Olkin (1985) described how the homogeneity analysis can be carried out using the General Linear Model program of the Statistical Analysis System (Statistical Analysis System Institute, 1985). The weighted statement is used to minimize the weighted residual sums of squares. In such an analysis, the sums of squares due to the tested moderator variable is treated as a chi-square statistic.

Study coding and intercoder reliability. Thirty-six different characteristics of each effect size were coded, including characteristics of the research report (e.g., author, journal, year), and characteristics of the judges, senders, and messages. Each sample was independently coded by two coders and any discrepancies were resolved in conference. Some of the coded variables did not prove useful (e.g., in only one study was the ethnicity of the judges reported), so they will not be described further.

The accuracy–confidence estimates. Most important to this report were the studies that reported correlations between confidence and accuracy in detecting deception. A total of 18 independent samples drawn from 16 research reports resulted in 27 correlation coefficients describing the relation between accuracy and confidence. Of the 16 reports, 15 were journal articles and one was an unpublished manuscript (see Table 1). The median year of publication was 1991. The number of comparisons per sample ranged from one (in 11 samples) to four (in 2 samples). The other charac-

Table 1. *Correlations Between Accuracy and Confidence*

Authors	Year	r
Anderson, Ansfield, & DePaulo[a]	1997	.04
Brandt, Miller, & Hocking	1980	−.01
DePaulo & Pfeifer[b]	1986	.02
Ekman & O'Sullivan[c]	1991	.03
Fan, Wagner, & Manstead	1995	−.11
Feeley & deTurck	1996	.07
Frank & Ekman	1997	
Study 1		.05
Study 2		.17
Kohnken	1987	.06
Lavrakas & Maier	1979	.20
Levine & McCornack	1992	−.14
Littlepage & Pineault	1981	.11
McCornack & Parks	1986	−.20
Smith, Archer, & Costanzo[d]	1991	
Study 1		.08
Study 2		.26
Swann, Silvera, & Proske	1995	.00
Vrij	1993	−.15
Vrij	1994	.04

[a]Mean of the correlations for two different points in time, −.02 and .09. [b]For the college students, new recruits, and advanced officers, the between-participants correlations were .14, .10, and −.12, respectively. The corresponding within-participants correlations were .03, .02, and .01. [c]There were 7 samples in this report but separate accuracy–confidence correlations were not reported for each one. The correlation reported here is the mean of pre- and postconfidence measures. [d]Only a subset of the items involved deception detection. However, those items were not analyzed separately from the others.

teristics reported next (of the judges, senders, etc.) are based on these 18 samples.

Judge characteristics. All of the judges in the 18 samples were adults. In 12 of the samples, the judges were exclusively or almost exclusively college students. (A few included high school students or graduate students, and one included some staff members in addition to the students.) In three samples (Kohnken, 1987; Vrij, 1993, 1994), all of the judges were police detectives. In DePaulo & Pfeifer (1986), there were two samples of federal law enforcement officers (new recruits and advanced officers) and one sample of college students. In Ekman & O'Sullivan (1991), there was one sample of college students, and six other groups: Secret Service, federal polygraphers, robbery investigators, judges, psychiatrists, and special interest groups (accountants, social workers, etc.). For the five studies that reported age (DePaulo & Pfeifer, 1986; Kohnken, 1987; McCornack & Parks, 1986; Vrij, 1993, 1994), the mean age was 29.8 years. (Because most of the studies involving college students as judges did not report age, the mean age across all of the samples was probably lower.) For one of the 18 samples (Brandt, Miller, & Hocking, 1980), the sex of the judges was unspecified, and in one other (Kohnken, 1987), all of the judges were men. All other samples included both men and women. However, when the judges were police officers or detectives (DePaulo & Pfeifer's 1986 new recruits and advanced officers; Ekman & O'Sullivan's 1991 robbery investigators; Vrij, 1993, 1994), the samples included many more men than women. In Ekman and O'Sullivan (1991), the samples of Secret Service, polygraphers, judges, and psychiatrists were also almost all men.

Sender characteristics. The senders were strangers to the judges in 15 of the samples. In one sample the sender was a same-sex friend (Anderson, Ansfield, & DePaulo, 1997), and in two the senders were dating the judges (Levine & McCornack, 1992; McCornack & Parks, 1986).

For the majority of samples, the senders were both men and women. The total number of female senders ranged from 2 to 45 and male senders also ranged from 2 to 45. Two samples included only men as senders (Frank & Ekman, 1997), and one sample included only women (Ekman & O'Sullivan, 1991).

Message and paradigm characteristics. The contents of the truths and lies were coded as fact based (12 samples), feeling based (5 samples), personal information (4 samples), and attitudinal information (7 samples). These categories were not mutually exclusive, and in seven samples, the messages were of more than one kind.

In nine samples, the messages were presented in the full audiovisual modality (face, body, and speech). (In two of those—Anderson et al., 1997, and Swann, Silvera, & Proske, 1995—the senders and judges interacted live.) Two samples included speech only (DePaulo & Pfeifer, 1986; Lavrakas & Maier, 1979), two included face plus speech (Frank & Ekman, 1997, Study 1, Study 2), and one had both visual and speech (with no indication of whether both face and body was included; Brandt et al., 1980). The modality by which the messages were presented was manipulated in three samples (Littlepage & Pineault, 1981; Smith, Archer, & Costanzo, 1991, Study 2; Vrij, 1994).

The length of the messages ranged from 44 s to 240 s in those samples that reported message length. The number of messages included in each sample ranged from 4 to 64. The number of truths equaled the number of lies in all 15 samples for which this information was reported.

Two of the samples used an interaction paradigm in which the senders described their truths and lies with no one else present (Kohnken, 1987; Lavrakas & Maier, 1979). In 13 samples the paradigms were partly interactive (i.e., an interviewer was present asking questions of the sender) and 2 were totally interactive (i.e., the sender and another person without a script interacted face-to-face; Anderson et al., 1997; Levine & McCornack, 1992; McCornack & Parks, 1986; Swann et al., 1995). In Smith et al. (1991), the degree of interaction varied across messages.

Characteristics of the measures. In two samples (DePaulo & Pfeifer, 1986; Smith et al., 1991, Study 2), the accuracy–confidence correlation was computed within-participants. In these samples, judges indicated their judgments of deceptiveness and their confidence for every message that they observed. Next, the two sets of ratings were correlated separately for each judge and then averaged across judges. In 17 samples, the correlations were between-participants. The judges' confidence ratings were averaged before they were correlated with their overall accuracy at detecting deception, or the judges provided only one report of their overall confidence—which was correlated with their overall accuracy. One study reported both within-participants and between-participants correlations (DePaulo & Pfeifer, 1986). Additionally, one research report contained two studies, one which reported a within-participants correlation and one which reported a between-participants correlation (Smith et al., 1991).

In all 18 samples the measure of confidence was continuous. In 14 samples, judges' overall accuracy at detecting deception was reported. Three reported accuracy rates less than the chance level of 50% (Brandt et al., 1980; Kohnken, 1987; Vrij, 1993) and the other 11 reported percentages greater than 50% (DePaulo &

Pfeifer, 1986; Ekman & O'Sullivan, 1991; Fan, Wagner, & Manstead, 1995; Feeley & deTurck, 1996; Frank & Ekman, 1997, Study 1, Study 2; Lavrakas & Maier, 1979; Littlepage & Pineault, 1981; McCornack & Parks, 1986; Swann et al., 1995; Vrij, 1994).

Results

Accuracy and Confidence

Overall relation. Table 2 presents a stem and leaf display of the correlations between confidence and accuracy at detecting deception from 18 independent samples. As shown in the display, the correlations ranged from –.20 to .26. The unweighted average relation between accuracy and confidence, based on these 18 correlations, was $r = .03$. When correlations were weighted by sample size, the overall average correlation was $r = .04$. The 95% confidence interval for the weighted average correlation ranged from .00 to .08 and thus included $r = .00$. Therefore, the null hypothesis, that no relation exists between confidence and accuracy in detecting deception, cannot be rejected.

Table 2. *Stem and Leaf Plot of Accuracy–Confidence Correlation Coefficients*

Stem	Leaf
+.2	06
+.1	17
+.0	023445678
–.0	1
–.1	145
–.2	0

Moderators of the accuracy–confidence relation. A homogeneity analysis revealed that there was no more variance in the set of 18 correlations than would be expected from sample error alone, $\chi^2(16) = 25.20$, $p > .05$. However, based on theoretical considerations, we continued our analysis to look for moderators of differences in effect sizes. For example, we tested the prediction made by Smith and her colleagues (Smith et al., 1991) that when judges make a series of judgments that are scored for accuracy and then make one rating of overall confidence at the end, the accuracy–confidence relation will be smaller than when they make a confidence judgment corresponding to each of their deceptiveness judgments. As shown in Table 3, there were no significant moderators of the accuracy–confidence relation.

Overconfidence

In six studies, confidence and accuracy were either measured on the same scales, or were converted to the same 100-point scale; and the overall means were compared (Brandt et al., 1980; Costanzo, 1992; deTurck, 1991; deTurck, Harszlak, Bodhorn, & Texter, 1990; McCornack & Parks, 1986; Swann et al., 1995). The

Table 3. *Moderators of the Accuracy–Confidence Relation*

Moderators	*n*	95% Confidence Interval		
		Low Estimate	*M*	High Estimate
Overall	18	.00	.04	.08
Aggregation $\chi^2(1) = 1.14$, $p > .05$[a]				
Between Participants	16	–.01	.03	.07
Within Participants	5	–.01	.07	.15
Were the Judges Law Enforcement Officers? $\chi^2(1) = 1.48$, $p > .05$[b]				
Officers	5	–.06	.01	.08
Students	14	.01	.06	.11
Modality $\chi^2(1) = 2.73$, $p > .05$				
Face, Body, and Speech	8	–.09	–.03	.03
All Other Modalities	5	–.03	.05	.13
Paradigm $\chi^2(1) = 2.90$, $p > .05$[c]				
No Interaction	2	–.01	.14	.29
Partly or Totally Interactive	14	–.04	.00	.04
Accuracy at Detecting Deception $\chi^2(1) = .16$, $p > .05$[d]				
Less Than 50% Accurate	3	–.17	–.04	.09
More Than 50% Accurate	11	–.02	.03	–.08

[a]DePaulo and Pfeifer (1986) contributed both between-participants and within-participants correlations to this comparison. The within-participants correlations were computed separately for the students, new recruits, and advanced officers. [b]DePaulo and Pfeifer (1986) contributed one correlation from students and two from officers. Ekman and O'Sullivan (1991) was removed from this analysis because they did not report separate *r*s for the student and law enforcement samples. [c]Smith, Archer, and Costanzo (1991) was removed from this analysis because the degree of interaction was different for different messages. [d]The statistic is calculated using accuracy as a continuous variable; however, the average *r* and confidence interval are presented for accuracy rates below and above 50%.

results indicated that judges' confidence ($M = 72.91\%$) was substantially higher than their accuracy ($M = 57.20\%$) in all of the studies.

Other Correlates of Confidence

Truth bias. As shown in Table 4, we were able to calculate nine effect sizes from eight studies that estimated the relation between truth bias and confidence (Anderson et al., 1997, which contributed two effect sizes; Fan et al., 1995; Kohnken, 1987; Levine & McCornack, 1992; Levine, McCornack, & Aleman, 1996; Littlepage & Pineault, 1981; McCornack & Parks, 1986; Vrij, 1994). The average unweighted correlation was $r = .16$. The weighted average correlation was $r = .17$, and the 95% confidence interval did not include $r = .00$ (.11, .23). This correlation indicated that judges who were more confident in their judgments

Table 4. *Other Correlates of Confidence*

Variable	Authors	Year	*r*
Truthfulness Bias			
	Anderson, Ansfield, & DePaulo	1997	
	Time 1		.25
	Time 2		–.16
	Fan, Wagner, & Manstead	1995	.27
	Kohnken	1987	.17
	Levine & McCornack	1992	.43
	Levine et al.	1996	.30
	Littlepage & Pineault	1981	.03
	McCornack & Parks	1986	.41
	Vrij	1994	–.38
	Overall weighted *r*		**.17**
	Overall unweighted *r*		**.16**
Relationship Development and Closeness			
	Anderson et al.	1997	
	"Feel close" Time 1		–.02
	"Feel close" Time 2		.40
	"Behave close" Time 1		–.01
	"Behave close" Time 2		.34
	Fleming, Darley, Hilton, & Kojetin[a]	1990	
	Study 2		.00
	Study 3		.00
	Levine & McCornack	1992	.34
	McCornack & Parks	1986	.26
	Overall weighted *r*		**.19**
	Overall unweighted *r*		**.16**
Sex of Judge			
	DePaulo & Pfeifer	1986	.11
	deTurck	1991	
	Trained men vs. trained women		.24
	Untrained men vs. untrained women		.02
	deTurck, Harszlak, Bodhorn, & Texter	1990	.12
	Smith, Archer, & Costanzo Study 1	1991	.12
	Toris & DePaulo	1984	.55
	Overall weighted *r*		**.15**
	Overall unweighted *r*		**.21**
Actual Truthfulness of the Messages			
	Allen & Atkinson	1981	–.55
	Anderson et al.	1997	.42
	Fan et al.	1995	–.18
	Fleming et al.	1990	
	Study 2		.25
	Study 3		.18
	Kohnken	1987	.00
	Littlepage & Pineault	1979	–.35
	Waxer	1983	.60
	Overall weighted *r*		**.15**
	Overall unweighted *r*		**.05**

Note: Positive correlations indicate that confidence was greater when truth bias was greater, when relationships were closer, when judges were men, and when judges were rating truths rather than lies.
[a]Estimates are extremely conservative.

were also more likely to perceive the messages as truths than as lies.

Relationship between sender and judge. Effect sizes estimating the link between closeness and confidence could be calculated from five independent samples (Anderson et al., 1997; Fleming, Darley, Hilton, & Kojetin, 1990, which contributed two effect sizes; Levine & McCornack, 1992; McCornack & Parks, 1986). The unweighted correlation was $r = .16$. The weighted average was $r = .19$. The 95% confidence interval did not include 0 (.10, .28). Thus, confidence in judgments of deception increased with the closeness of the relationship (see Table 4).

Also, the closeness of the relationship between the sender and the judge was a significant moderator of the relation between truth bias and confidence, $\chi^2(1) = 18.70$, $p < .05$. Correlations between confidence and truth bias were larger for studies in which the senders and judges were friends or were dating, $r = .38$, than for studies in which the senders and judges did not know one another, $r = .08$.

Finally, there were seven studies that contained information regarding the relationship between the senders and judges and also the percentage of truths guessed by the judges. In the studies in which the senders and judges were dating partners or friends (Anderson et al., 1997; McCornack & Parks, 1986), the judges guessed truth 66% of the time. In the studies in which the judges and senders were strangers (Fan et al., 1995; Kohnken, 1987; Littlepage & Pineault, 1981; Swann et al., 1995; Vrij, 1993), the judges guessed truth 59.38% of the time.

Sex of judge. There were five studies that reported enough information to calculate an effect size indicating whether men (coded as 1) and women (coded as −1) differed in their confidence in their deceptiveness judgments (DePaulo & Pfeifer, 1986; deTurck, 1991; deTurck et al., 1990; Smith et al., 1991; Toris & DePaulo, 1984). The correlation was positive in every study (see Table 4). The unweighted average correlation was $r = .21$. The weighted average correlation was $r = .15$, with a 95% confidence interval that did not include .00 (.09, .21). Thus, men were significantly more confident about their judgments of deception than women.

One additional study that did not contain enough information to calculate an accuracy–confidence effect size but did include comparisons between the sexes also indicated that men were more confident than women (Hurd & Noller, 1988). In that study, confidence was assessed from judges' tone of voice as they discussed their decisions rather than from self-reports.

Actual truthfulness of the messages. Eight independent samples examined whether judges' confidence differed when they were rating truths compared to when they were rating lies. The average unweighted correlation was $r = .05$. The weighted average correlation was $r = .15$. The 95% confidence interval did not include 0 (.06, .24). Thus, judges were significantly more confident when judging truths than when judging lies (Allen & Atkinson, 1981; Anderson et al., 1997; Fan et al., 1995; Fleming et al., 1990; Kohnken, 1987; Littlepage & Pineault, 1979; Waxer, 1983). In another study in which confidence was assessed from judges' tone of voice as they discussed their judgments about whether the message they just heard was a truth or a lie, the judges sounded significantly more confident when the message they were discussing really was a truth than when it was a lie (Hurd & Noller, 1988).

Discussion

Accuracy and Confidence

As predicted, people's confidence in their judgments of whether another person is telling the truth or lying is not significantly related to the accuracy of those judgments. Twenty-seven correlations from 18 samples yielded a mean weighted correlation of only .04. The largest reported correlation was .26. There were no significant moderators of the size of the accuracy–confidence correlation. The estimate of a near-zero correlation is based on a total of 2,972 judges, including law enforcement officers as well as students, and friends and lovers of the senders as well as strangers.

Methodological explanations. Measurement considerations might be used to explain the independence of accuracy and confidence in detecting deception. That is, the measures of either accuracy or confidence might lack construct validity. This appears not to be the case.

With regard to accuracy at detecting deception, the argument might be made that because accuracy hovers around chance in many studies it cannot be expected to correlate substantially with any other variables. However, research indicates that accuracy at detecting deceit, though not impressive, is generally better than chance (DePaulo et al., 1985). This result was replicated in the present review: Accuracy was better than chance (though not always significantly so) in 11 of the 14 samples that reported overall accuracy scores. Further, accuracy measures have been shown to be related in predictable ways to factors such as age (DePaulo, Jordan, Irvine, & Laser, 1982), personality (DePaulo & Tang, 1994), and culture (Bond, Omar, Mahmoud, & Bonser, 1990).

Of course, it is still possible that the correlation between accuracy and confidence was higher for stud-

ies in which overall accuracy at detecting deceit was greater than chance than for those in which it was less than chance. Our results did not support that possibility (see Table 3).

The results of the present review also indicated that measures of confidence in deception judgments, although unrelated to accuracy, are related to other variables in theoretically meaningful ways. As we discuss next, the accumulated research indicates that confidence is related to the judge's bias to perceive messages as truths, the relationship between the sender and the judge, the sex of the judge, and the actual truthfulness of the messages. Thus, the trustworthiness of the measures appears not to explain the lack of an accuracy–confidence relation.

Cognitive explanations. An important reason why confidence is unrelated to accuracy may be that people's theories about cues to deception are in some ways erroneous (DePaulo et al., 1985; Fiedler & Walka, 1993; Zuckerman, Koestner, et al., 1981). The cues that people believe to be indicative of deceit sometimes are not, and cues that are in fact valid are sometimes overlooked. When judges notice cues that they mistakenly believe to be associated with deception, or which they believe to be associated with deception more reliably than they are in fact, their confidence will exceed their accuracy (Griffin & Tversky, 1992)—and the accuracy–confidence link will be severed.

There also are indications that people's theories about cues to deception are insufficiently differentiated. For example, people seem to use similar cues to infer when men and women are lying, even though there are differences in the cues to men's and women's lies (DePaulo & Rosenthal, 1979; DePaulo, Rosenthal, Rosenkrantz, & Green, 1982).

More generally, accuracy and confidence may be produced by different kinds of mechanisms and therefore be related to different kinds of factors. For example, the judges' sex predicts confidence but not accuracy. In the studies covered in this review, men were more confident than women, but in the literature on deception in general (Zuckerman, DePaulo, et al., 1981), men are not more accurate detectors. Similarly, the closeness of the relationship between the sender and judge was positively correlated with confidence in the studies we reviewed. Yet, evidence is surprisingly weak for the intuitively appealing prediction that the closeness of the sender–judge relationship should also be positively correlated with accuracy (Anderson, Ansfield, & DePaulo, in press). Developmental changes provide another example. Accuracy increases with age from childhood to adulthood (DePaulo et al., 1982). However, it is not yet clear whether confidence increases with age as well. The overconfidence effect, whereby confidence is inappropriately high relative to accuracy, may also be consistent with a multiple mechanism model.

Overconfidence

Confidence is not just uncorrelated with accuracy, it is sometimes substantially greater than accuracy. In all six studies in which mean levels of confidence and accuracy at detecting deception could be compared, confidence was higher than accuracy. This overconfidence effect, which has been documented in other kinds of judgmental tasks of moderate to extreme difficulty (Fischhoff et al., 1977; Koriat et al., 1980), can be of great practical significance. Swann and his colleagues (1995) argued that people might use their own feelings of knowing their partner to decide whether sex with that partner would be safe; but if their feelings of confidence in their judgments are inappropriately high, then the decisions they make may be deadly.

Other Correlates of Confidence

Closeness and truth bias. We found support for several of the links in the McCornack and Parks (1986) model of deception detection in personal relationships. The model predicts that as relationships develop and become closer, partners will become more confident in their judgments of each other's truths and lies. Consistent with this prediction, we found that relationship closeness was positively correlated with confidence.

The second step in their model predicts that as confidence increases, so too will the tendency to perceive messages as truths, regardless of the actual truthfulness of the messages. In support of this prediction, we found that confidence and truth bias were positively correlated. We also found that the relation between feelings of confidence and judgments of truthfulness is stronger when the judge is friends with the sender or dating the sender than when the sender and judge are strangers.

We also examined the link between closeness and truth bias and found that when the judges were friends with the senders or dating them, they were more likely to guess that the messages were truths than when they were strangers. According to the McCornack and Parks (1986) model, this link between closeness and truth bias is an indirect one that is mediated by confidence. If confidence is partialled out of the relationship, the correlation should no longer be significant. In their own work, McCornack and his colleagues presented evidence consistent with this argument (Levine & McCornack, 1992; McCornack & Parks, 1986). We have not found any other evidence to add to theirs.

Although the results of our review underscore the promise of the McCornack model, questions about

causality still remain. Methodologically, it might be most convincing if the key variables could be experimentally manipulated. For example, if strangers could be induced to feel emotionally closer to each other, would they then feel more confident in their judgments of each other's truths and lies than strangers who did not experience the emotional induction? If judges could be randomly assigned to conditions in which their confidence was or was not bolstered, would the more confident judges show a greater truth bias? It might be argued that the model pertains to actual relationships and the confidence that develops within those relationships, and not to short-term experimental manipulations of the variables. In that case, causality could be more convincingly established by longitudinal research.

In cognitive models of confidence in human judgments, such as Griffin and Tversky's (1992), the key theoretical constructs are aspects of the judgmental task (such as the strength and validity of the evidence at hand). Theoretically, an important implication of the link between relationship closeness and confidence is that confidence may follow from variables that are extraneous to the deception-detection task. The finding that men have more confidence in their deception-detection judgments than do women is another example of this.

Confidence might also follow from social roles or from theories and expectations about the characteristics of people in different social roles. For example, DePaulo & Pfeifer (1986) compared the detection skills of undergraduates, new recruits to federal law enforcement jobs, and advanced federal law enforcement officers. They found that both groups of officers were more confident than the college students, and the new recruits who had been on the job an average of only 5 months were just as confident as the officers who had an average of more than 7 years of experience. Perhaps even the effects of judges' sex and relationship status are theory based: People may believe that they should be better at detecting the deception of their close relationship partners, and men may believe that they should be better than women at detecting deception.

To the extent that the officers in the DePaulo and Pfeifer (1986) study were more confident than the students because they thought that, as officers, they should be especially skilled at detecting deceit, the results extend the Griffin and Tversky (1992) model by suggesting the potential importance of factors extraneous to the judgmental task. However, the results can also be construed as consistent with Griffin and Tversky's (1992) prediction that in domains in which predictability is low (as in the detection of deception), experts may evidence even greater overconfidence effects than novices. The very rich models of the domains that are developed by experts lead them to feel more confident in their judgments than do novices, who do not have such complex theories to bolster their confidence (see also Cantor & Kihlstrom, 1987; Kruglanski, 1989).

Actual truthfulness of the messages. We also tested the hypothesis that feelings of confidence might function as measures of indirect deception detection, in that they might differentiate truths from lies. The meta-analysis of eight tests of this hypothesis produced support for this idea. Judges were substantially more confident when judging truths than when judging lies. Additional supportive evidence came from a study in which confidence was not based on judges' self-reports but on their tone of voice as they talked through the process of deciding whether the message they just heard was a truth or a lie (Hurd & Noller, 1988).

Measures of indirect deception detection hold great promise in this field in which explicit measures of deception detection often yield unimpressive levels of accuracy. Judges who appear to be totally unable to distinguish truths from lies based on their explicit judgments may show some evidence of accurate discrimination based on indirect measures (DePaulo et al., 1982). Such judges may not realize that the discriminations they are making are relevant to deception detection, or they may simply be unwilling explicitly to call another person a liar.

Conclusion

In the literature on eyewitness accuracy, early reports that accuracy and confidence were usually unrelated (e.g., Deffenbacher, 1980; Wells & Murray, 1984) were followed by waves of research showing that the correlation was substantial under certain theoretically predictable conditions—such as when information-processing conditions were optimal (Deffenbacher, 1980) or when the witnesses could watch a tape of their own testimony before indicating their confidence (Kassin, 1985; Kassin, Rigby, & Castillo, 1991).

Will future research on detecting deception similarly point to more optimistic possibilities? As we indicated in the introduction, the potential for substantial positive correlations is more limited in the deception-detection domain. In the realm of eyewitness testimony, the accuracy–confidence correlation might be enhanced by conditions that strengthen the memory trace—for example, seeing the target person's face for longer periods of time (Bothwell et al., 1987). In the area of detecting deception, a stronger memory for a message may not add to the accuracy–confidence link if the judges' theories about cues to deception are wrong.

Although the confidence of particular individuals may be unrelated to their accuracy, as long as the correlation is not exactly zero, a substantial relation could exist at an aggregate level. As Smith et al. (1991)

documented, an impressive correlation (.74) can be generated when accuracy scores and confidence ratings for a given message are averaged across many judges. The implication is that messages that generally are more easily read are also more confidently read. However, although this finding is of some interest theoretically, it is of little practical use to the individual who wants to know if it is wise to use one's own confidence in one's judgments as a guide to the accuracy of those judgments. The answer to that question is still no.

References

Allen, V. L., & Atkinson, M. L. (1981). Identification of spontaneous and deliberate behavior. *Journal of Nonverbal Behavior, 5,* 224–237.

American Psycological Association. (1974–present). *PsycLIT* [CD-ROM]. Wellesley Hills, MA: SilverPlatter Information Services [Producer and Distributor].

Anderson, D. E., Ansfield, M. E., & DePaulo, B. M. (1997). *Accuracy and confidence in the detection of deception in developing friendships.* Unpublished data, University of Virginia, Charlottesville.

Anderson, D. E., Ansfield, M. E., & DePaulo, B. M. (in press). Love's best habit: Deception in the context of relationships. In P. Philippot, R. S. Feldman, & E. J. Coats (Eds.), *The social context of nonverbal behavior.* Cambridge, England: Cambridge University Press.

Bond, C. F., Omar, A., Mahmoud, A., & Bonser, R. N. (1990). Lie detection across cultures. *Journal of Nonverbal Behavior, 14,* 189–204.

Bothwell, R. K., Deffenbacher, K. A., & Brigham, J. C. (1987). Correlation of eyewitness accuracy and confidence: Optimality hypothesis revisited. *Journal of Applied Psychology, 72,* 691–695.

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

Cantor, N., & Kihlstrom, J. F. (1987). *Personality and social intelligence.* Englewood Cliffs, NJ: Prentice-Hall.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Cooper, H. M. (1989). *Integrating research: A guide for literature reviews.* Newbury Park, CA: Sage.

Cooper, H. M., & Hedges, L. V. (1994). *The handbook of research synthesis.* New York: Russell Sage Foundation.

Costanzo, M. (1992). Training students to decode verbal and nonverbal cues: Effects on confidence and performance. *Journal of Educational Psychology, 84,* 308–313.

Deffenbacher, K. A. (1980). Eyewitness accuracy and confidence: Can we infer anything about their relationship? *Law and Human Behavior, 4,* 243–260.

DePaulo, B. M. (1994). Spotting lies: Can humans learn to do better? *Current Directions in Psychological Science, 3,* 83–86.

DePaulo, B. M., Jordan, A., Irvine, A., & Laser, P. S. (1982). Age changes in the detection of deception. *Child Development, 53,* 701–709.

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., & Rosenthal, R. (1979). Telling lies. *Journal of Personality and Social Psychology, 37,* 1712–1722.

DePaulo, B. M., Rosenthal, R., Rosenkrantz, J., & Green, C. R. (1982). Actual and perceived cues to deception: A closer look at speech. *Basic and Applied Social Psychology, 3,* 291–312.

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

DePaulo, B. M., & Tang, J. (1994). Social anxiety and social judgment: The example of detecting deception. *Journal of Research in Personality, 28,* 142–153.

deTurck, M. A. (1991). Training observers to detect spontaneous deception: Effects of gender. *Communication Reports, 4,* 81–89.

deTurck, M. A., Harszlak, J. J., Bodhorn, D. J., & Texter, L. A. (1990). The effects of training social perceivers to detect deception from behavioral cues. *Communication Quarterly, 38,* 189–199.

Ekman, P., & O'Sullivan, M. (1991). Who can catch a liar? *American Psychologist, 46,* 913–920.

Fan, R. M., Wagner, H. L., & Manstead, A. S. R. (1995). Anchoring, familiarity, and confidence in the detection of deception. *Basic and Applied Social Psychology, 17,* 83–96.

Feeley, T. H., deTurck, M. A. (1996, April). *Baseline relevance and global cue priming in lie detection.* Paper presented at the Eastern Communication Association, New York.

Fiedler, K., & Walka, I. (1993). Training lie detectors to use nonverbal cues instead of global heuristics. *Human Communication Research, 20,* 199–223.

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1977). Knowing with certainty: The appropriateness of extreme confidence. *Journal of Experimental Psychology: Human Perception and Performance, 3,* 552–564.

Fleming, J. H., Darley, J. M., Hilton, J. L., & Kojetin, B. A. (1990). Multiple audience problem: A strategic communication perspective on social perception. *Journal of Personality and Social Psychology, 38,* 593–609.

Frank, M. G., & Ekman, P. (1997). The ability to detect deceit generalizes across different types of high-stake lies. *Journal of Personality and Social Psychology, 72,* 1429–1439.

Griffin, D., & Tversky, A. (1992). The weighing of evidence and the determinants of confidence. *Cognitive Psychology, 24,* 411–435.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* New York: Academic.

Hurd, K., & Noller, P. (1988). Decoding deception: A look at the process. *Journal of Nonverbal Behavior, 12,* 217–233.

Ickes, W. (1993). Empathic accuracy. *Journal of Personality, 61,* 587–610.

Kassin, S. M. (1985). Eyewitness identification: Retrospective self-awareness and the accuracy–confidence correlation. *Journal of Personality and Social Psychology, 49,* 878–893.

Kassin, S. M., Rigby, S., & Castillo, S. R. (1991). The accuracy–confidence correlation in eyewitness testimony: Limits and extensions of the retrospective self-awareness effect. *Journal of Personality and Social Psychology, 61,* 698–707.

Kohnken, G. (1987). Training police officers to detect deceptive eyewitness statements: Does it work? *Social Behavior, 2,* 1–17.

Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 107–118.

Kraut, R. E., & Poe, D. (1980). Behavioral roots of person perception: The deception judgments of customs inspectors and laymen. *Journal of Personality and Social Psychology, 39,* 784–798.

Kruglanski, A. W. (1989). The psychology of being "right": The problem of accuracy in social perception and cognition. *Psychological Bulletin, 106,* 395–409.

Lavrakas, P. J., & Maier, R. A. (1979). Differences in human ability to judge veracity from the audio medium. *Journal of Research in Personality, 13,* 139–153.

Levine, T. R., & McCornack, S. A. (1992). Linking love and lies: A formal test of the McCornack and Parks model of deception detection. *Journal of Social and Personal Relationships, 9,* 143–154.

Levine, T. R., McCornack, S. A., & Aleman, C. G. (1996). *Ask and you shall believe: Behavioral adaptation, confidence, and heuristic-based explanations of the probing effect.* Unpublished manuscript, University of Hawaii, Manoa.

Lichtenstein, S., & Fischhoff, B. (1977). Do those who know more also know more about how much they know? The calibration of probability judgments. *Organizational Behavior and Human Performance, 20,* 159–183.

Littlepage, G. E., & Pineault, M. A. (1979). Detection of deceptive factual statements from the body and the face. *Personality and Social Psychology Bulletin, 5,* 325–328.

Littlepage, G. E., & Pineault, M. A. (1981). Detection of truthful and deceptive interpersonal communication across information transmission modes. *Journal of Social Psychology, 114,* 57–68.

McCornack, S. A., & Parks, M. R. (1986). Deception detection and relationship development: The other side of trust. In M. L. McLaughlin (Ed.), *Communication yearbook, 9* (pp. 377–389). Beverly Hills: Sage.

Statistical Analysis System Institute. (1985). *SAS user's guide: Statistics* (Version 5 ed.). Cary, NC: Author.

Smith, H. J., Archer, D., & Costanzo, M. (1991). "Just a hunch": Accuracy and awareness in person perception. *Journal of Nonverbal Behavior, 15,* 3–18.

Swann, W. B., Jr., Silvera, D. H., & Proske, C. U. (1995). On "knowing your partner": Dangerous illusions in the age of AIDS? *Personal Relationships, 2,* 173–186.

Thomas, G., Fletcher, G. J. O., & Lange, C. (1997). On-line empathic accuracy in marital interaction. *Journal of Personality and Social Psychology, 72,* 839–850.

Toris, C., & DePaulo, B. M. (1984). Effects of actual deception and suspiciousness of deception on interpersonal perceptions. *Journal of Personality and Social Psychology, 47,* 1063–1073.

Vrij, A. (1993). Credibility judgements of detectives: The impact of nonverbal behavior, social skills, and physical characteristics on impression formation. *Journal of Social Psychology, 133,* 601–610.

Vrij, A. (1994). The impact of information and setting on detection of deception by police detectives. *Journal of Nonverbal Behavior, 18,* 117–136.

Waxer, P. H. (1983). Emotional deceit: False words versus false action. *Motivation and Emotion, 7,* 365–376.

Wells, G. L., & Murray, D. M. (1984). Eyewitness confidence. In G. Wells & E. Loftus (Eds.), *Eyewitness testimony: Psychological perspectives* (pp. 155–170). Cambridge, England: Cambridge University Press.

Zuckerman, M., DePaulo, B. M., & Rosenthal, R. (1981). Verbal and nonverbal communication of deception. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 14, pp. 1–59). New York: Academic.

Zuckerman, M., Koestner, R., & Driver, R. (1981). Beliefs about cues associated with deception. *Journal of Nonverbal Behavior, 6,* 105–114.