Accuracy in Detecting Truths and Lies: Documenting the "Veracity Effect"

Timothy R. Levine, Hee Sun Park and Steven A. McCornack

Deception research has consistently shown that accuracy rates tend to be just over fifty percent when accuracy rates are averaged across truthful and deceptive messages and when an equal number of truths and lies are judged. Breaking accuracy rates down by truths and lies, however, leads to a radically different conclusion. Across three studies, a large and consistent veracity effect was evident. Truths are most often correctly identified as honest, but errors predominate when lies are judged. Truth accuracy is substantially greater than chance, but the detection of lies was often significantly below chance. Also, consistent with the veracity effect, altering the truth-lie base rate affected accuracy. Accuracy was a positive linear function of the ratio of truthful messages to total messages. The results show that this veracity effect stems from a truth-bias, and suggest that the single best predictor of detection accuracy may be the veracity of message being judged. The internal consistency and parallelism of overall accuracy scores are also questioned. These findings challenge widely held conclusions about human accuracy in deception detection. **Key words:** Veracity Effect, Deception, Lie Detection

Many studies have examined human accuracy in detecting deception, and there is a strong general agreement among researchers about human's relative inability to detect deception in others. Most deception scholars agree that accuracy rates tend to be just over fifty percent. The findings supporting this conclusion, however, most often involve summing across truthful and deceptive messages where there is an equal distribution of truths and lies (Miller & Stiff, 1993). We argue that because people are most often truth-biased, message veracity is an important and often ignored determinant of detection accuracy. Specifically, we expect detection accuracy to be substantially above fifty percent for truths, but well below fifty percent for lies. We refer to this anticipated finding as "the veracity effect."

Detection Accuracy

Consensus among deception scholars is that people's ability to distinguish truths from lies tends to be significantly, but only slightly, better than chance levels. Across studies, meta-analysis indicates that the mean accuracy rate is about 57% (Kraut, 1980), and literature reviews conclude that the accuracy rates reported in individual studies almost always fall within the range of 45% to 70% (e.g., Kalbfleisch, 1994; Miller and Stiff, 1993). Even under the most advantageous conditions, observed detection rates are below 75% (see DeTurck, Harszlak, Bodhorn & Texter, 1990, for an exception). For example, individuals who receive detection training (e.g., deTurk & Miller, 1990; Fieldler and Walka, 1993), have considerable professional experience (e.g., Burgoon, Buller, Ebesu & Rockwell, 1994; Ekman & O'Sullivan, 1991; Kraut & Poe, 1980) or have access to idiosyncratic baseline information (e.g., Feeley,

Timothy R. Levine (Ph.D., Michigan State University, 1992) is Associate Professor in the Department of Speech at the University of Hawaii at Manoa. Hee Sun Park (M.A., University of Hawaii, 1998) is doctoral student in Communication at University of California Santa Barbara. Steven A. McCornack (Ph.D., University of Illinois, 1990) is Associate Professor in the Department of Communication at Michigan State University. *The authors appreciate the help of Dr. Michael Ogden in editing the tapes used in Study IV.* *Direct all correspondence to: Tim Levine, Department of Speech, 2560 Campus Road, University of Hawaii at Manoa, Honolulu, HI 96822, e-mail: mit@hawaii.edu.*

deTurck, & Young, 1995) also fall victim to this well documented accuracy ceiling. In short, the belief that deception detection accuracy rates are only slightly better than fifty-fifty is among the most well documented and commonly held conclusions in deception research. Our informal review of the recent literature indicates that these conclusions are echoed by many (e.g., Anderson, Ansfield & DePaulo, 1997; Burgoon et al., 1994; DePaulo, Kirkendol, Tang, & O'Brien, 1988; Feeley et al., 1995; Kalbfleisch, 1994; Millar & Millar, 1995; Miller & Stiff, 1993; O'Sullivan, Ekman & Friesen, 1988; Stiff, Kim & Ramesh, 1992; Stiff & Miller, 1986; Stiff et al., 1989; Toris & DePaulo, 1985; Vrij, 1994) and disputed by few (see Buller & Burgoon, 1996, p. 228 for an apparent exception).[1]

There are several reasonable explanations for peoples' relatively poor performance in deception detection. First, while a variety of verbal and nonverbal correlates of deception have been isolated (for reviews, see Burgoon, Buller & Woodall, 1989; Miller & Stiff, 1993; Zuckerman, DePaulo & Rosenthal, 1981), there is no behavior or set of behaviors that infallibly distinguishes deception from truth-telling (Kraut, 1980). Therefore, perfectly accurate deception detection is currently not possible.

Second, research-naive people seem to focus on the wrong behaviors when trying to distinguish truths from lies (Miller & Stiff, 1993; Stiff & Miller, 1986). Although no "sure-fire" deception cues exist, some statistically reliable correlates of deception have been isolated (Kraut, 1980; Zuckerman et al., 1981), as have several behaviors that research-naive receivers tend rely upon when making deception judgments (e.g., Stiff & Miller, 1986; Zuckerman et al.). When comparing the "authentic" deception cues to the behaviors that people tend to use, it becomes obvious that people are often influenced by some behaviors that lack predictive utility and people often ignore other diagnostically useful behaviors (Buller, Stryzewski & Hunsaker, 1991; Fieldler & Walka, 1993; Miller & Stiff, 1993; Stiff et al., 1989; Stiff & Miller, 1986).

Third, and most importantly for the current investigation, research indicates that people's veracity judgments are often affected by a variety of systematic errors and biases. Instead of actively scrutinizing another's behaviors for clues to deceit, receivers may instead rely on relatively mindless decision rules often labeled cognitive heuristics. Some of the heuristics that are believed to influence veracity judgments include the availability heuristic (O'Sullivan et al., 1988), the falsifiability heuristic (Fiedler & Walka, 1993), the infrequency heuristic (Fiedler & Walka, 1993), the probing heuristic (Levine & McCornack, 1994), the relational truth-bias heuristic (Stiff et al., 1992), and the representativeness heuristic (Stiff et al., 1989). Such heuristics may produce systematic errors in judgements and consequently account, in part, for poor accuracy rates.

The most well documented and commonly accepted of these biases is the so-called truth-bias (McCornack & Parks, 1986; originally labeled the "truthfulness bias" by Zuckerman et al., 1981 and Zuckerman, Koestner, Colella & Alton, 1984). Truth-bias refers to the tendency to judge more messages as truths than lies (Anderson, Ansfield, & DePaulo, 1997). Numerous studies have found that independent of actual message veracity, individuals are much more likely to ascribe truth to other's messages than deceit (e.g., Buller & Burgoon, 1996; Buller et al., 1991b; Burgoon et al., 1994; McCornack & Levine, 1990; O'Sullivan et al., 1988; Stiff et al., 1992; Zuckerman et al., 1981; Zuckerman et al., 1984). This bias is particularly strong among relational partners (Millar & Millar, 1995) and face-to-face conversational participants (Buller et al., 1991b), and holds true even under conditions of high contextual suspicion (McCornack & Levine, 1990). Truth-bias may stem from how people mentally represent information (Gilbert, Krull & Malone, 1990), from the fundamental inferential processes that guide conversational understanding (McCornack, 1992), face maintenance concerns (Anderson et al., 1997), and/or from a reliance on cognitive heuristics (Stiff et al., 1992).

Lie Accuracy vs. Truth Accuracy

In their recent review, Miller and Stiff (1993) observed that, "as used in the literature, deception detection is something of a misnomer" (p. 68). Typical procedures expose judges to several messages, half of which are truthful and half of which are deceptive. Judges are asked to distinguish the truths from the lies, and deception detection accuracy is calculated as the average accuracy across truths and lies. To avoid semantic confusion, Miller and Stiff (1993) advocate using "detection accuracy" for combined scores and reserving "deception detection" for the accuracy of detecting lies. Although Miller and Stiff carefully limit their comments to the semantic implications of this distinction, we believe this distinction is of potentially great methodological and substantive importance.

As shown above, that a) accuracy rates are, on average, only slightly better fifty percent, and that b) people exhibit a rather strong and persistent truth-bias are perhaps the two most widely accepted and well documented findings in deception research. Yet, one might question if these two conclusions are entirely consistent with one another. The existence of a strong truth-bias suggests that the veracity of the message judged may be an important determinant of detection accuracy. To the extent that people are truth-biased, and to the extent that chance or guessing contributes to accuracy rates, laws of probability dictate that people are more likely to correctly identify truths than lies (Zuckerman et al., 1984). Thus, accuracy rates for truths should be considerably higher than accuracy rates for lies. If this is the case, conclusions regarding accuracy rates may be an artifact of how detection accuracy is calculated, and breaking accuracy rates down by truths and lies may yield dramatically different conclusions.

Although relatively few studies have reported separate accuracy rates (Miller & Stiff, 1993), many that have seem to suggest a difference. In their meta-analysis, Zuckerman et al. (1981) examined 15 studies and concluded that "truthful messages were detected with more accuracy than deceptive ones" (p. 24). More recently, Zuckerman et al. (1984), Stiff and Miller (1986), and Millar and Millar (1997) have reported large effects for message veracity on detection accuracy (r = .54, .61, and .87 respectively) where truthful messages were evaluated more accurately than deceptive messages. Feeley et al. (1995) report significant differences in same direction. Finally, the message veracity effect has also been documented in interactive studies. For example, Burgoon et al. (1994) report that "interviewers were less accurate judging deception than they were judging truth" (p. 318), and Buller et al., (1991b) observed 81% accuracy for truthtellers but only 29% accuracy for liars.

Some conflicting claims and results, however, are evident. In one of the earliest studies of detection accuracy, Fay and Middleton (1941) found that truths were correctly identified 51% of the time while lies were detected with 61% accuracy. Ekman and Friesen's (1974) results indicated a nonsignificant trend toward lies being judged more accurately than truths. Similarly, Fieldler and Walka (1993) report nearly equal levels of accuracy (truth accuracy = 58% and lie accuracy = 62%) with lies judged slightly more accurately than truths. More recently still. Buller and Burgoon (1996a; 1996b; Burgoon, Buller, Ebesu, White & Rockwell, 1996; Burgoon, Buller, Dillman & Walther, 1995; Burgoon, Buller, Guerrero, Afifi & Feldman, 1996) argue that people are in fact accurate at detecting lies. Burgoon et al. claim to have "confirmed (as have others) that people sense deception when it is present" (1995, p. 184) and in summarizing their extensive line of research, Buller and Burgoon note "several of our studies have documented that receivers do perceive deception when it is present" (1996, p. 228).

Even when considering the exceptions noted above, the literature suggests message veracity is often a strong and reasonably consistent predictor of detection accuracy. Based on the strong empirical evidence consistent with a truth-bias and the preponderance of previous research, we hypothesize a "*veracity effect.*" Specifically, *individuals will be significantly more accurate at judging truths than lies* (H1).

If hypothesis one is true, the commonly held view that accuracy rates hover around 57% would be somewhat misleading in that accuracy would depend upon whether one was judging truths or lies. If an equal number of true and deceptive messages are judged, and if hypothesis one is consistent with the data, we further predict that *for honest statements, accuracy will be significantly greater than chance* (H2). Alternatively, *accuracy will be significantly lower than chance rates for deceptive messages* (H3).

Analyzing truth and lie accuracy separately has implications beyond general conclusions about distinguishing honesty from deceit. Specifically, distinguishing truth accuracy from lie accuracy may alter the relationships between a variety of other variables and accuracy. As a case in point, consider Brandt's often cited studies (Brandt, Miller, & Hocking, 1980a; 1980b; 1982) on familiarity and accuracy. These studies treat familiarity as a varying number of pre-exposures to honest messages, and report an approximately linear increase accuracy through three prior exposures. These findings have become commonly accepted in the literature (e.g., Comadena, 1982; Levine & McCornack; 1992; McCornack & Parks, 1986; Miller & Stiff, 1993). In Feeley et al.'s (1995) recent replication of these findings, however, the linear familiarity effects were confined to honest messages. Familiarity did not affect lie detection. Hence, one might wonder if message veracity moderates other accuracy findings.

Study 1

Our initial examination of the suspected veracity effect involves a reanalysis of McCornack and Levine's (1990) previously published data on suspicion and detection accuracy among romantic partners. Several studies have failed to find significant main effects for suspicion on accuracy (Burgoon et al., 1994; Stiff et al., 1992; Toris & DePaulo, 1984). McCornack and Levine attribute these null findings to a nonlinear effect (inverted U) for suspicion on accuracy, and found apparent support for the nonlinear explanation. McCornack and Levine's data were deemed appropriate for the current investigation because a) separate accuracy results were not provided in the original report, b) the design allowed for a test of the three hypotheses presented above, and c) the study provided an example of how examining separate accuracy results may alter the substantive conclusions drawn from previously published data.

Method

Participants and Procedures

The participants were 104 heterosexual romantically-involved couples recruited from communication classes at a large Midwestern University. Experimental sessions were run separately for each couple. The participants ranged in age from 18 to 36 ($M = 19.8$) and had been dating for an average of 22.89 months. One member was randomly selected to play the role of "source" and the other was selected to play the role of "receiver."

The source was taken to a separate room to create a videotape containing a series of truthful and deceptive statements. First, the source completed an attitude survey containing 12 randomly selected Machiavellianism scale items (Christie & Geis, 1970). Each item used a 10-point response format. After completing each of the 12 items, the experimenter changed half of the items to an answer which was five points different (half the scale length) from the subjects true answer. The confederate was then told to report his or her true answer for half of the 12 items. For the remaining (changed) items, the confederate was told to report the answer which was five points (half the scale length) different from his or her true answer. The ordering was varied so that each item was reported truthfully or dishonestly an equal number of times over the course of the entire study. The source was then seated at a table and told to state his or her attitude on each item in terms of the true or false scale value, and to briefly explain each answer. This was videotaped. The camera was placed so that the head, arms, and torso of the confederate would appear.

Upon completion, the tape was taken to their partner (i.e., the "receiver") who viewed it. Receivers were randomly assigned to one of three suspicion conditions. In the low suspicion condition, no information was given to the receiver regarding the potential for deceptiveness on the part of their partner. The experimenter referred to all items as items designed to "assess truthfulness", and the truth-lie questions were imbedded within 4 filler items. Respondents in the moderate suspicion condition were told by the experimenter prior to completing the questionnaire that there was the possibility that their partner may not be completely truthful in how they answer each item. In the high suspicion condition, receivers were informed that their partner would definitely be lying on several of the items, and that their task was to determine which of their partner's responses were deceptive.

After viewing each of the 12 videotaped segments, subjects were asked to make several judgments including a dichotomous "truth/lie" judgment which served as a measure of both truth-bias and accuracy. Judgmental accuracy at differentiating truths and lies were computed as the proportion of accurate judgments to total judgments for the 6 truths and lies respectively (0–100% accuracy). Truth-bias was calculated as the proportion of total items judged as truthful.

A one-way ANOVA was conducted to ascertain if a manipulation check scale varied as a function of the experimental condition. The results indicated a significant and substantial effect for the experimental condition upon the manipulation check items [$F(2, 95) = 10.86, p < .0001, \eta^2 = .19, r = .44$). The obtained cell means were ordered in the predicted direction (low = 22.36, moderate = 25.97, and high = 34.03), and when the sum of squares were decomposed the linear component accounted for 94% of the explained sum of squares. A more complete description of the method is reported in McCornack and Levine (1990).

TABLE 1
EFFECTS OF AROUSED SUSPICION ON DETECTION ACCURACY OF TRUTHS AND LIES

Statement	Aroused Suspicion: Low	Moderate	High	Across Conditions
Truths	84.3%	87.9%	73.5%	82.8%
Lies	22.5%	37.4%	43.1%	34.3%

Results and Discussion

Consistent with previous studies, overall accuracy was slightly, but significantly above chance levels, $M = 58\%$, $SD = 13\%$, $t(99) = 8.00$, $p < .0001$, $r = .63$. Also consistent with previous research, a substantial truth-bias was evident (74% of all messages were judged true, $SD = 15\%$). The proportion of truth judgments was significantly greater than the actual distribution of truths and lies, $t(99) = 24.00$, $p < .0001$, $r = .92$, indicating a strong truth-bias.

A 2 (truth, lie) × 3 (low, moderate, and high suspicion) mixed ANOVA indicated that each component variation was statistically significant. Consistent with hypothesis one, a substantial main effect was found for veracity; $F(1, 98) = 326.21$, $p < .0001$, $\eta^2 = .60$, $r = .77$. As expected, truth accuracy ($M = 82\%$, $SD = 17\%$, 95% confidence interval, 78% to 86%) was greater than lie accuracy ($M = 34\%$, $SD = 22\%$, 95% confidence interval, 30% to 38%). Also as predicted, truth accuracy was significantly greater than chance [$t(102) = 16.00$, $p < .0001$, $r = .85$] while lie accuracy was significantly lower than chance, $t(102) = -8.00$, $p < .0001$, $r = -.62$.

The main effect for suspicion was also significant [$F(2, 98) = 4.50$, $p < .02$, $\eta^2 = .01$, $r = .12$], but qualified by a larger suspicion by veracity interaction [$F(2, 98) = 12.26$, $p < .001$, $\eta^2 = .05$, $r = .21$]. Examination of the means in table 1 shows that while the nonlinear main effect reported by McConack and Levine (1990) held for truthful messages [$t(101) = 2.88$, $p < .002$, $r = .28$, 49% of explained sums of squares] and total accuracy [$t(101) = 2.53$, $p < .02$, $r = .24$, 77% of explained sums of squares], it was not evident for lie accuracy. Instead, a linear trend accounting for 95% of explained sums of squares was observed for lies, $t(101) = 4.98$, $p < .001$, $r = .44$.

A few supplemental correlation analyses were conducted. As might be expected, truth bias was positively related to truth accuracy [$r(99) = .66$, $p < .0001$] and negatively associated with lie accuracy [$r(99) = -.82$, $p < .0001$]. Truth accuracy and lie accuracy were not correlated, $r(99) = -.12$, $p = .11$.

Thus, the data were consistent with each of the three hypotheses. Even though overall accuracy (58%) was virtually identical to the across-study average (57%, Kraut, 1980), truth accuracy (82%) was more than twice as high as lie accuracy (34%) and the effect size for veracity ($\eta^2 = .60$) overwhelmed all other systematic and random effects. Also, truth accuracy was significantly greater than chance while lie accuracy was significantly lower than chance. Clearly then, claims regarding unaided human detection accuracy depend upon whether the person is assessing a truth or lie.

These data also provide another example of how calculating separate accuracy rates can alter substantive conclusions regarding other predictors of accuracy. Here, McCornack and Levine's (1990) conclusions about the nonlinear effects of suspicion were confined to truthful messages and it appears that their published conclusions are an artifact of summing across truths and lies.

Although the data were consistent with our three hypotheses, the preceding comments should be interpreted with caution. Foremost among the qualifications necessary is possibility of a confirmation bias. Truth-baises are positively correlated with truth accuracy and negitively associated with lie accuracy. Further, relational partners are believed by some to be more truth biased than strangers. Therefore, the use of romantically involved couples may have created an experimental environment favorable to the predications advanced, and replication with strangers is necessary. Studies II and III provide such replications with previously unpublished data.

Study II

The second study replicates the first with two major differences. First, message sources were recruited separately from the judges making relational involvement unlikely. Second, probing was manipulated instead of suspicion. Previous research indicated that probing is not related to overall accuracy (Buller, Comstock, Aune & Strzyzewski, 1989; Buller, Strzyzewski & Comstock, 1991; Stiff & Miller, 1986), but is positively associated with truth-bias (Buller et al., 1989; Buller, Strzyzewski & Comstock, 1991). Because probing may alter source behaviors (Buller et al., 1989; Buller, Stiff & Burgoon, 1996; Buller, Strzyzewski & Comstock, 1991; Stiff & Miller, 1986) which may have additional effects on veracity judgments, source behaviors were held constant and confounding was avoided.

Method

Participants and Procedures

Seventeen undergraduate students (11 females, 6 males) from a large mid-western university served as message sources. Each completed a four-item questionnaire containing randomly-selected items from the Mach IV scale (Christie & Geis, 1970). Responses to these items were recorded on 10-point Likert-type scales. Two of the respondent's answers were changed to a value five points different from the respondent's original true answers. Six different orderings of changed items were used to minimize order effects.

Respondents were then informed that s/he would be interviewed regarding her/his answers on the altered questionnaire. The respondents were told to report her/his original answer for the two unchanged items during the interview. For the remaining two items, the respondents were told to report the answer which was changed. Each respondent was given 10 minutes to prepare her/his responses.

At the end of this time, an experimenter entered the room, and sat facing the respondent. Each respondent was asked to state her/his attitude on each of the items, and to briefly explain her/his answer. The experimenter was not informed as to which items were truthful (i.e., unchanged) and which items were lies (i.e., changed). This interview was videotaped from behind a one-way mirror. The camera was placed so that the respondent's entire body would appear, while the experimenter could not be seen.

During the interview, each respondent reported her/his answer and then briefly discussed reasons for her/his choice of that answer. Following each answer, the experimenter probed the respondent with a neutral probe, saying "tell me a little bit more about why you answered it that way." Following this probe, each respondent elaborated briefly upon her/his answer.

A single item was randomly selected as the basis for all responses on the tapes in order to minimize potential effects due to variation in item content. Eight videotaped respondents were then selected from the original sample of 17 interviewees: two males who had lied, two males who had told the truth, two female liars, and two female truthtellers. A master tape was made, with each of these eight interviews. The order of the interviews was randomly determined.

Three additional tapes were created from the master tape. In one, the audio-track of the tape was altered, so that the experimenter's original "neutral" probe was replaced with a negative, face-threatening probe. The experimenter said "I don't think you really circled that. Tell me a little more about why you answered it that way." In a second, the experimenter's original probe was replaced with a positive, face-supportive probe. The positive probe was "that sounds reasonable, but tell me a little more about why you answered it that way." In the third stimulus tape, the experimenter's probe was deleted and the tape spliced, so that each respondent's answer to the item and her/his explanation of the answer were joined together. This resulted in four tapes each corresponding to four probing conditions: no probe, neutral probe (i.e., the master tape), positive probe, and negative probe.

Three-hundred and thirty-seven participants (129 men and 208 women) were then solicited to judge one of the four tapes. Each experimental session involved having a group of five to seven participants view one of the four test tapes, which were randomly assigned. Fifty-seven respondents viewed the neutral probe tape, 100 respondents viewed the negative probe tape, 95 respondents viewed the positive probe tape, and 85 respondents viewed the no probe tape. After viewing each of the eight videotaped interviews, respondents were asked to rate their perceptions of the source's honesty on a single dichotomous "truth/lie" item. Perception of source honesty was computed as the proportion of truthfulness judgments to total judgments across the eight cases (0-100% honesty). Accuracy for truths and lies were also calculated as percentages.

Results and Discussion

Consistent with previous studies and study I, overall accuracy was slightly, but significantly [$t(328) = 3.00$, $p < .01$, $r = .19$] above chance levels ($M = 53\%$, $SD = 16\%$). Also consistent with previous research, a substantial truth-bias was evident ($M = 72\%$, $SD = 18\%$, of all messages were judged true). The frequency of truth judgments was significantly greater than the actual distribution of truths and lies, $t(328) = 22.00$, $p < .0001$, $r = .82$.

The data were analyzed with a 2 (truth, lie) × 4 (positive, negative, neutral, no probe) mixed ANOVA. Consistent with hypothesis one, a substantial main effect was found for veracity; $F(1,325) = 499.09$, $p < .0001$, $\eta^2 = .45$, $r = .67$. Truth accuracy ($M = 75\%$, $SD = 21\%$, confidence interval = 77% to 73%) was again greater than lie accuracy ($M = 31\%$, $SD = 26\%$, confidence interval = 29% to 33%). Also as predicted, truth accuracy was significantly greater than chance [$t(330) = 25.00$, $p < .0001$, $r = .81$] while lie accuracy was significantly lower than chance, $t(330) = -19.00$, $p < .0001$, $r = -.72$.

The main effect for probing was not significant [$F(3, 325) = 1.37$, $p = ns$, $\eta^2 = .00$,

TABLE 2

EFFECTS OF PROBING ON DETECTION ACCURACY OF TRUTHS AND LIES

Statement	Neutral	Neg	Pos	None	Across Conditions
Truths	75.0%	76.0%	78.7%	70.0%	75.2%
Lies	23.7%	31.6%	29.7%	38.1%	31.3%

$r = .05$], but the probing by veracity interaction [$F(3, 325) = 4.67, p < .003, \eta^2 = .01, r = .11$] was significant. Examination of the means in table 2 shows that lies were detected with the most accuracy, and truths were judged with the least accuracy in the no probe condition.

The correlation analyses reveal strikingly similar results to those of study one. Truth bias was positively related to truth accuracy [$r(328) = .68, p < .0001$] and negatively associated with lie accuracy [$r(328) = -.80, p < .0001$]. Truth accuracy and lie accuracy were negatively correlated, $r(328) = -.10, p < .03$.

The data were once again consistent with each of the three hypotheses. Truth accuracy (75%) was again more than twice as high as lie accuracy (31%) and the effect size ($\eta^2 = .45$) was again large. Similarly, lie accuracy was significantly less than chance while truth accuracy was significantly greater than chance. Therefore, the veracity effect is not confined to romantic partners.

Study II also provided yet another example of how calculating separate accuracy rates can affect the interpretation of results. Previous studies have found that probing is not related detection accuracy (Stiff & Miller, 1986; Buller et al., 1989; Buller et al., 1991). When breaking accuracy down by message veracity, however, significant effects are apparent. Probing appears to improve truth accuracy but inhibits lie accuracy.

Study III

The results thus far have been quite consistent and lead to three related conclusions. First, the judges in both studies exhibited a substantial truth bias ($M = 74\%$, $r = .92$ and 72%, $r = .82$ respectively). Second, truth-bias was positively associated with truth detection accuracy (rs = .66 and .68) and negatively related to lie accuracy (rs = −.82 and −.80). Third, and apparently as a consequence, truthful statements were identified with greater accuracy than deceptive statements (rs = .77 and .67). Hence, the veracity effect seems to be a function of truth-bias. Study III further explores the link between truth-bias and accuracy.

As noted earlier, truth-bias may result, in part, from heuristic processing (Stiff et al., 1992). If this is the case, factors that affect heuristic processing may also affect the veracity effect, albeit indirectly. Specifically, when judging lies under conditions in which heuristic processing is more likely, subjects should be more truth-biased and lies should be detected with less accuracy. Alternatively, when heuristic processing is less likely, truth biases should decline, and lie accuracy should increase as a result.

This reasoning, however, may not hold for truth accuracy. While a decline in heuristic processing should make the chances of *guessing* a truth correctly less likely, any decline in heuristic processing should also reduce the reliance on guessing as a strategy for truth detection. Therefore, a reduced reliance on heuristic processing in judging honest messages need not produce a decline in the number of truth judgements or in truth accuracy. Instead, decreased heuristic processing may increase truth judgments and truth accuracy as a function of more deliberate and unerring processing.

Stiff et al. (1989) suggested that the heuristic processing of deceptive messages is determined, in part, by an individual's familiarity with the information presented in the message. Individuals are able to process messages involving familiar information more actively than those involving unfamiliar information (Stiff et al., 1989). Thus, relative to individuals confronted with familiar information, individuals faced with messages involving unfamiliar information should be more likely to rely upon heuristics in rendering veracity judgments.

The above reasoning leads to additional hypotheses regarding the veracity effect, truth-bias, and familiarity with message content. Topic familiarity should reduce the reliance on heuristic processing, and as a result, the number of truth judgements should decline for lies but not for truths. More precisely stated, *content familiarity will interact with message veracity to affect truth bias such that for lies, truth bias will be greater for unfamiliar messages than familiar messages, but for truthful messages, truth bias will be greater for familiar messages than unfamiliar messages* (H4). The net result should be that *detection accuracy for both truths and lies will be greater under conditions of content familiarity than under conditions of unfamiliarity* (H5).

Because a main effect for familiarity on accuracy is expected in hypothesis 5, hypotheses 1 remains unaltered. Honest statements should be judged more accurately in both familiar and unfamiliar conditions. Hypothesis 2, predicting that truth accuracy is above chance levels, should be exacerbated by content familiarity. That is, truth accuracy should be even higher when message content is familiar. Hypothesis 3, however, may depend on heuristic processing and thus might only apply to unfamiliar content. Specifically, lie accuracy may be significantly below chance levels with unfamiliar, but not familiar message content.

Method

Participants and Procedures

Two undergraduate students (one female, one male) from a large mid-western university served as message sources. The sources completed an eight-item questionnaire on issues thought to be controversial on college campuses. Four of the items involved controversies specific to the campus at which the experiment was being conducted, while the remaining four items dealt with issues specific to a different campus out-of-state. Responses to these items were recorded on 10-point Likert-type scales. After completing the eight items, an experimental assistant collected the questionnaire, and changed four of the respondent's answers to answers which were five points different (i.e., half of the scale length) from the respondent's original true answers. Four different orderings of which items were to be changed were used to minimize order effects.

Each respondent was then informed that s/he would be interviewed regarding her/his answers on the altered questionnaire. Each respondent was told to report her/his original answer for the four unchanged items during the interview. For the remaining four items, each respondent was told to report the answer which was altered by the assistant. Each respondent was given 10 minutes to prepare her/his responses. At the end of this time, the experimenter entered the room, and sat facing the respondent. Each respondent was asked to state her/his attitude on each of the items, and to briefly explain her/his answer. The experimenter was not informed as to which items were truthful (i.e., unchanged) and which items were false (i.e., changed). This interview was videotaped. The camera was placed so that the respondent's entire body would appear, while the experimenter could not be seen. The camera was placed behind a one-way mirror.

During the interview, each respondent reported her/his answer and then briefly discussed reasons for her/his choice of that answer. Following each answer, the experimenter probed the respondent with a neutral probe, saying "tell me a little bit more about why you answered it that way." Following this probe, each respondent elaborated briefly upon her/his answer.

In creating the familiar content tape, the four items (for each source) involving familiar topics were preserved, and the items involving off-campus information were deleted. Thus, the tape was comprised of eight responses total: four familiar items discussed by the female source (two truthful answers, two false answers), and four familiar items discussed by the male source (two truthful, two false). A similar procedure was then followed in the construction of the unfamiliar content tape, so that it involved the discussion of eight unfamiliar items (four discussed by the female source, four discussed by the male source).

One-hundred and thirty-six respondents (80 women and 56 men) were solicited to serve as judges. Each experimental session involved having a group of five to seven judges view one of the four test tapes. Each group was randomly assigned to view one of the two test tapes previously generated (i.e., familiar or unfamiliar). Sixty-six respondents viewed the familiar tape, and 70 respondents viewed the unfamiliar tape.

After viewing each of the eight videotaped interviews on the test tape to which they had been assigned, the tape was stopped, and respondents were asked to indicate if they thought that the individual on the tape was "lying" or "completely truthful." Truth-bias and accuracy were computed as percentages. Following completion of the eighth judgment, respondents were asked to complete a familiarity manipulation-check items scale.

The effectiveness of the manipulation of familiarity was tested using a four-item measure designed by the experimenters. Sample items included "I was familiar with the issues that the subjects on this videotape discussed" and "I had no prior knowledge of the topics that were discussed by the subjects on this videotape." The manipulation check items were tested for unidimensionality with confirmatory factor analysis. All four manipulation check items were retained ($M = 18.73$, $SD = 7.94$, $\alpha = .93$). A one-way ANOVA was used to assess the quality of the familiarity manipulation. Subjects rated the content of the messages as significantly more familiar ($M = 25.81$) in the condition involving the discussion of familiar items than in the condition involving unfamiliar items [$M = 12.64$; $F(1, 134) = 225.92$, $p < .0001$, $\eta^2 = .63$, $r = .79$] suggesting that the manipulation was effective.

Results and Discussion

Again, overall accuracy was slightly, but significantly [$t(134) = 4.00$, $p < .001$, $r = .33$] above chance levels ($M = 58\%$, $SD = 28\%$). Also, the proportion of truth judgments significantly exceeded the actual distribution of truths and lies [$t(134) = 12.43$, $p < .0001$, $r = .73$] indicating a substantial truth-bias ($M = 69\%$, $SD = 18\%$).

TABLE 3

EFFECTS OF FAMILIARITY ON THE DETECTION ACCURACY OF TRUTHS AND LIES

Statement	Familiar	Unfamiliar	Across Conditions
Truths	85.4%	68.6%	76.7%
Lies	52.5%	26.7%	39.2%
Across Conditions	69.0%	47.7%	58.0%

TABLE 4

EFFECTS OF FAMILIARITY AND MESSAGE VERACITY ON TRUTH-BIAS

Statement	Familiar	Unfamiliar	Across Conditions
Truths	85.4%	68.6%	76.7%
Lies	47.5%	73.3%	60.8%

The accuracy data were analyzed with a 2 (truth, lie) × 2 (familiar, unfamiliar) mixed ANOVA. Consistent with hypothesis one, a substantial main effect was found for veracity; $F(1, 134) = 118.95$, $p < .0001$, $\eta^2 = .31$, $r = .55$. Truth accuracy ($M = 77\%$, $SD = 23\%$, confidence interval = 81% to 73%) was again greater than lie accuracy ($M = 39\%$, $SD = 32\%$, confidence interval = 33% to 45%). Also as predicted, truth accuracy was significantly greater than chance [$t(134) = 13.50$, $p < .0001$, $r = .76$] while lie accuracy was significantly lower than chance, $t(134) = -3.67$, $p < .01$, $r = -.30$.

As predicted in hypothesis 5, the main effect for familiarity was significant [$F(1, 134) = 55.45$, $p < .0001$, $\eta^2 = .10$, $r = .31$], and was not qualified by a familiarity by veracity interaction [$F(1, 134) = 1.75$, $p = ns$. $\eta^2 = .00$, $r = .07$]. Examination of the means in table 3 shows that across truths and lies, accuracy was higher in the familiar condition (69%) than in the unfamiliar condition (48%).

The truth bias predictions made in hypothesis 4 were tested with a similar 2 (truth, lie) × 2 (familiar, unfamiliar) mixed ANOVA. Consistent with hypothesis 4, the veracity by familiarity interaction was significant, $F(1, 134) = 55.45$, $p < .0001$, $\eta^2 = .13$, $r = .36$. A weaker main effect for veracity was also evident ($F(1, 134) = 33.44$, $p < .0001$, $\eta^2 = .08$, $r = .28$]. The familiarity main effect was nonsignificant, $F(1, 134) = 1.75$, $p = ns$, $\eta^2 = .00$, $r = .07$. As can be seen in Table 4, subjects were substantially truth biased in all conditions except the familiar-lie cell. Consistent with hypothesis 4, an analysis of simple effects showed that for lies, the number of truth judgments was significantly greater in the unfamiliar condition (73%) than in the familiar condition (47%), $F(1, 134) = 25.53$, $p < .0001$, $\eta^2 = .16$, $r = .40$. The reverse was the case for truthful messages. The number of truth judgments was significantly greater in the familiar condition (85%) than in the unfamiliar condition (69%), $F(1, 134) = 20.59$, $p < .0001$, $\eta^2 = .13$, $r = .37$. Thus, as heuristic processing becomes less likely, the proportion of truth judgments decrease for lies and increase for truths.

The supplemental correlations were consistent with those reported previously. Truth bias was positively related to truth accuracy [$r(134) = .50$, $p < .0001$] and negatively associated with lie accuracy [$r(134) = -.75$, $p < .0001$]. Truth accuracy and lie accuracy, were uncorrelated, $r(134) = -.01$, $p = ns$.

The data were once again consistent with each of the three original hypotheses. Truth accuracy (77%) was almost twice as high as lie accuracy (39%) and the effect size ($\eta^2 = .31$) was again large. Familiarity with message content also had a substantial effect on accuracy, increasing truth accuracy by 16.8% and lie accuracy by 25.8%. This main effect is consistent with Stiff et al.'s, (1989) contention that content familiarity does enhance one's ability to actively assess message veracity thereby reducing the reliance on heuristic processing. But, because content familiarity did not interact with message veracity, the veracity effect holds under conditions where heuristic processing should be less likely. This finding is consistent with McCornack's (1997) argument that truth-bias may, in part stem from processes more fundamental than heuristic processing.

Overall lie accuracy was significantly less than chance while truth accuracy was significantly greater than chance. As expected, however, these conclusions must be qualified by content familiarity. Lie accuracy was below chance levels in the unfamiliar condition (27%), but not in the familiar condition (53%).

Study IV

In their review, Miller and Stiff (1993) observed the typical procedures used in detection accuracy studies expose judges to several messages, half of which are truthful and half of which are deceptive. The existence of a veracity effect suggests that the ratio of honest messages to the total number of messages judged (i.e., truth-lie base rate) may be an important determinant of detection accuracy. Because accuracy rates for truths are considerably higher than accuracy rates for lies, previous conclusions regarding accuracy may be an artifact of the common truth-lie base rate (50% truth) used in the literature. Different base rates should result in different accuracy rates.

To our knowledge, only one study has used a base-rate of other than 50% honest. Geizer, Rarick & Soldow (1977) employed a design in which two-thirds of the messages judged were deceptive. The accuracy rate in that study was 40.3%, slightly above chance levels, but well below the accuracy levels of studies using a 50% base-rate.

Based on the strong empirical evidence consistent with veracity effect, we hypothesize a variation that might be called a base-rate effect. Specifically, *truth-lie base rates will significantly affect detection accuracy such that accuracy will be the highest in a 75% honest condition, significantly lower in a 50% honest condition, and lowest in a 25% honest condition* (H6). Further, we anticipate that *accuracy will be a linear function of the base rate* (H7).

Method

Stimulus Materials and Experimental Design

Two undergraduate students (one male and one female) from an upper division Speech class at the University of Hawaii served as message sources in exchange for extra credit. A videotape containing a series of truthful and deceptive statements was made for each source. First, the source completed an attitude survey containing 12 Machiavellianism scale items (Christie & Geis, 1970). Each item used a 10-point response format. After completing each of the 12 items, the experimenter changed half of the items (selected at random) to an answer which was five points different (half the scale length) from the subject's true answer. The source was then told to report his or her true answer for half of the 12 items. For the remaining (changed) items, the source was told to report the answer which was five points (half the scale length) different from his or her true answer. The source was then seated in a chair and told to state his or her response to each item, and to briefly explain each answer. This was videotaped. The camera was placed so that the source's entire body was visible. This procedure resulted in the creation of two (one male source and one female source) "master tapes," each containing 12 statements, 6 true and 6 lies.

Each master tape was digitized and edited with an AVID computer video editing system. First, 3 second pauses were placed between each of the twelve segments. Second, four segments were deleted from each tape. In the 75% honest condition, four of the six lies were deleted. In the 25% honest condition, four of the six honest answers were deleted. Two honest statements and two lies were removed to form a 50% honest condition.

Thus, this procedure resulted in the creation of 6 stimulus tapes, each with eight segments separated by 3 second pauses. The two message sources were crossed with the three base-rate (75%, 50%, and 25% honest) creating two tapes for each of the three experimental conditions. The two tapes for each condition pictured a different source and a different ordering of truths and lies. In every case, whether an item was a truth or lie, and whether or not an item was deleted was determined by random.

Participants and Procedures

The participants were 177 (90 males and 87 females) undergraduate students enrolled in an introductory Speech class at the University of Hawaii at Manoa. The participants were ethnically diverse (30.5% Japanese, 16.9% Chinese, 13.6% Filipino, 11.9% Caucasian, 7.9% Mixed, 7.3% Hawaiian, 5.6% other, 2.8% Korean, 2.3% African American, 0.6% Hispanic, and 0.6% Samoan,), and ranged in age from 16 to 38 ($M = 20.82$, $SD = 3.64$). All received extra credit in exchange for their participation.

Experimental sessions were run in groups of seven or fewer outside regular class time. Each group was randomly assigned to view one of the six stimulus tapes, with the constraint that an approximately equal number of participants viewed each tape. Between 28 and 30 participants viewed each tape (N = 58, 60, and 59 in the 75%, 50%, and 25% honest conditions respectively; N = 90 and 87 for the male and female source respectively).

The respondents were told that the study concerned people's perception of others, and that they would be watching a series of videotaped segments in which another student was discussing his or her answers at an attitude survey. The participants were told that students sometimes do not answer surveys honestly, and that the experimenter wanted to know if they thought the source was answering honestly or not. After the instructions were completed, the participants were shown each video clip one at a time. After viewing each of the 8 videotaped segments, the tape was paused, and the subjects were asked to make a dichotomous "truth/lie" judgment which served as a measure of both truth-bias and accuracy. Judgmental accuracy at differentiating truths from lies was computed as the proportion of accurate judgments to total judgments across the 8 segments (0-100% accuracy). Truth-bias was calculated as the percentage of total items judged as truthful. The participants also completed demographic items and were debriefed.

Results

Prior to testing the hypotheses, the tapes from the different sources were tested for additivity. The source of the tape did not affect accuracy [main effect, $F(1, 171) = 0.32$, $p = ns$, $\eta^2 = .00$, interaction with experimental condition, $F(2, 171) = 2.40$, $p = ns$, $\eta^2 = .03$]. Therefore, scores were summed across the two sources with each base-rate condition.

Hypothesis 6 predicted that the truth-lie base rates would significantly affect detection accuracy such that accuracy will be the highest in a 75% honest condition, significantly lower in a 50% honest condition, and lowest in a 25% honest condition. The data were consistent with this prediction. The main effect for base-rate was statistically significant and substantial, $F(2, 171) = 22.92$, $p < .0001$, $\eta^2 = .21$. All means were in the predicted direction, and all three means differed significantly from one another with conservative Scheff'e and Bonferroni procedures. Accuracy was 59.48% (SD = 14.49%) in the 75% honest condition, 51.88% (SD = 16.5%) in the 50% honest condition, and 39.83% (SD = 16.65%) in the 25% honest condition.

Hypothesis 7 predicted linear effects. The sums of squares were decomposed, and the linear component was significant, $F(1, 174) = 45.05$, $p < .0001$, $\eta^2 = .21$. The linear effect explained 98.3% of the explained sums of squares indicating an excellent fit to the data.

An exploratory analysis examined the effects of base-rate and message source on truth-bias. Truth-bias did not vary as a function of base-rate, $F(2, 171) = 1.67$, $p = ns$, $\eta^2 = .02$, and a truth-bias was evident across conditions (M = 68%, SD = 16.6%). The truth-bias means were 68% in the 75% honest condition, 70% in the 50% honest condition, and 65% in the 25% honest condition. The female source (71%) was seen as more honest than the male source [63%; $F(1, 171) = 11.76$, $p = .001$, $\eta^2 = .06$. Source and base-rate did not interact, $F(2, 171) = 1.30$, $p = ns$, $\eta^2 = .02$.

General Discussion

Perhaps the most frequently cited finding in deception research is that accuracy rates tend to be just over fifty percent. The findings leading to this conclusion, however, average accuracy rates across truthful and deceptive messages where an equal number of truths and lies are judged (Miller & Stiff, 1993). Breaking accuracy rates down by truths and lies and altering the ratio of truths to lies leads us to question this commonly held conclusion.

We argued that the veracity of the message under scrutiny is an important, and often ignored, determinant of detection accuracy. The veracity effect holds that truths are identified with greater accuracy than lies. Across the four studies, a strong and persistent veracity effect was evident. In the first three studies, people most often correctly identified truths as honest, but most often erred in identifying lies as deceptive. The margin of the difference between truth accuracy and lie accuracy was striking (47.5%, 43.9%, 37.5% respectively) as were the effect sizes ($r = .77$, .67, and .55 respectively). The veracity effect was found to hold for relational partners and strangers, whether or not subjects were primed to be suspicious, whether or not sources were asked probing questions, and whether or not message content was familiar. The consistency of these findings and the large effects reported leads us to conclude that *the single best predictor of detection accuracy is the veracity of message being judged.*

TABLE 5

CORRELATIONS AMONG TRUTH-BIAS, TRUTH ACCURACY AND LIE ACCURACY

Study	Truth-Bias and Truth Accuracy	Truth-Bias and Lie Accuracy	Truth and Lie Accuracy
Study I	−.66**	.82**	[illegible]
Study II	−.58**	.75**	[illegible]
Study III	.50**	[illegible]	.01

Note. *$p < .05$; **$p < .0001$.

TABLE 6

EFFECT SIZES (R) FOR TRUTH-BIAS AND THE VERACITY EFFECT

Study	Truth-Bias Effect	Veracity Effect
Study I	.92*	.77*
Study II	.82*	.67*
Study III	.73*	.55*

Note. *$p < .0001$.

Truth-bias has been advanced here and elsewhere (cf., Zuckerman et al., 1981; Zuckerman et al., 1984) as an explanation for the veracity effect. The data presented above document the relationship between truth-bias and the detection of truths and lies. Truth bias was positively correlated with truth accuracy ($r = +.50$ to $+.66$) and negatively correlated with lie accuracy ($r = -.75$ to $-.82$) in each of the first three studies (see table 5). That a truth-bias drives the veracity effect can be demonstrated even more dramatically by comparing the effect size for truth-bias obtained in first three studies with the corresponding effect size for the veracity main effect (see Table 6). When correlating the two effect sizes, a near perfect linear association is revealed $r(1) = .997, p < .05$.

Noting that the veracity effect results from a truth bias is not a sufficient explanation because this raises the question of why people are so truth-biased in the first place. There have been several explanations advanced for the truth-bias, and some of the most prominent ones are discussed here. Stiff et al. originally (1992) linked truth bias to heuristic processing, and this relationship was further explored in study III. Consistent with the heuristic explanation, people were more accurate in the familiar message conditions where heuristic processing should be less likely. However, because the veracity effect held across familiarity conditions, other explanations should be considered.

Based on a highly influential series of studies by Gilbert et al. (1990), McCornack (1997) concluded that "although heuristics undoubtedly influence deception judgments, there is reason to believe that humans' deficiency as lie detectors derives from more deeply-rooted cognitive processes related to knowledge representations" (p. 103). Gilbert et al.'s findings on the representation of information suggest that incoming information is, by default, considered truthful, and information considered false must be subsequently judged and relabeled as such (Gilbert et al., 1990). Such processing results in much more efficient processing of information, but this efficiency comes, at times, at the cost of accuracy.

Finally, Anderson et al. (1997) argue that truth-bias is probably an accurate reflection of everyday life. They believe that most of the messages people hear outside the deception laboratory are probably truthful. Consistent with this argument, DePaulo, Kashy, Kirkendol, Wyer and Epstein (1996) found that people report telling relatively few lies in everyday situations. Therefore, it may make sense for people to develop a tendency to judge others as honest. If this is correct, people are probably more accurate than previous research suggests. According to this view, because previous studies under represent the level of honesty in typical conversations, they also underestimate detection accuracy in everyday interactions.

Many deception researchers, however, argue that deception is a frequent occurrence. In an often cited study, Turner, Edgely and Olmstead (1975) reported that some form of information control was evident in the majority of conversational statements. If Turner et al.'s data are accurate, and if information control is equated with deception, then previous accuracy studies may have actually underestimated the frequency of deceit present in everyday conversation. If this is the case, previous studies would have underestimated detection accuracy and accuracy may be significantly below chance levels.

There is of course no definitive answer to the question of the actual base-rate in the typical everyday conversation. The likelihood of deceit surely varies dramatically across people and situations. Whatever the base-rate, however, the current results suggest that the single best predictor of detection accuracy is the veracity of the message under scrutiny.

These findings have several important implications for deception research. Most obviously, these findings challenge the widely accepted claim that detection accuracy is slightly better than chance. Average detection accuracy rates *cannot* be generalized to lie detection rates, and studies using a 50-50 base-rate *cannot* be generalized to other base rates. Simply put, most previous findings are artifactual, and lie accuracy is often significantly below chance levels.

Second, message veracity has been found to moderate the effects of several important predictors of detection accuracy. The effects of familiarity (Feeley et al., 1995), suspicion (study I), and probing (study II) on accuracy have failed to generalize across veracity conditions. Thus, the role of variables antecedent to detection accuracy must be reassessed in light of the current results.

Third, the very validity of calculating total accuracy scores is questioned. The units comprising valid measures must be internally consistent (the "items" should be positively correlated) and parallel (i.e., have similar correlations with outside measures). The findings from each of the first three studies indicate that neither criteria is met. Truth accuracy and lie accuracy are not positively correlated (see table 5), nor are they parallel with respect to suspicion or probing. Thus, calculating total accuracy scores is inappropriate.

Because beliefs in current conclusions about deception accuracy are so widely accepted and so seldom questioned, we anticipate many readers may be skeptical of our data and conclusions. Specifically, we anticipate three sets of objections.

First, some might dismiss our conclusions on the grounds that they are neither new nor original and hence are uninformative. Indeed, findings similar to ours have been around at least since Zuckerman et al.'s (1981) meta-analysis. Yet, although similar findings have been reported with some frequency, authors have uniformly down-played or ignored them. For example, Zucherman et al. (1981) dismissed the veracity effect as bias and imply that it is therefore uninteresting. Zucherman et al. (1984) relegate many of their comments about truth bias and accuracy to endnote (see p. 302). Although the veracity effect was by far the largest effect (η^2 = .37) reported by Stiff and Miller (1986), they ignored it completely in their discussion, instead devoting two pages to a much smaller probing effect (η^2 = .02). Similarly, Buller et al. (1991b) made little of their findings of 81% accuracy for truthtellers and 29% accuracy for liars, and more recently have even argued that research participants do detect deceit when present (Buller & Burgoon, 1996a; Buller et al., 1996; Burgoon et al., 1995, 1996).

Second, while some might grant that the veracity effect has been ignored, they would argue that this is because the effect is uninteresting or unimportant. Yet, as shown above, the veracity effect has several important substantive and methodological implications. Stated succinctly, the existence of the veracity effects suggests that current conclusions about detection accuracy may be the result of a methodical artifacts stemming from the use of total accuracy scores and an equal distribution of truths and lies.

Third, some might challenge the current conclusions on methodological grounds. Such critics might argue that because all four studies were noninteractive, all employed sanctioned lies, and all used dichotomous truth-lie judgments, our conclusions are suspect. We would adamantly disagree. First, there is reason to believe that the current findings do generalize to interactive deception and to unsanctioned lies. The veracity effect has been reported in interactive studies (Buller et al., 1991b; Burgoon et al., 1994), and with unsanctioned lies (Stiff and Miller, 1986). Further, the use of dichotomous measures was appropriate in our studies because these measures allowed for accuracy to be calculated as a percentage (allowing a comparison to previous findings), and because the messages being judged were dichotomous maintaining a direct correspondence between the stimulus and the measurement. Some have argued for the superiority of continuous truth-lie measures because these also tap uncertainty (e.g., Burgoon et al., 1995). We choose to measure certainty with separate items to avoid confounding (although we did not report our confidence findings).

In conclusion, the data from three studies were consistent with a veracity effect. Truths are judged with substantially greater accuracy than lies. These effects were replicated and the effect sizes were large. These findings have important implications for deception research and suggest that we may need to reassess many commonly held conclusions about deceptive communication.

End Notes

[1]Although Burgoon and Buller argue that people can detect deception when present, their results are consistent with those of other researchers and the findings of the present investigations. Burgoon and Buller have repeatedly found a) people rate honest messages significantly but only slightly more honest than lies, and b) both truths and lies tend to be rated as more honest than deceptive.

References

Anderson, D.E., Ansfield, M.E., & DePaulo, B.M. (1997). 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: Cambridge University Press.

Brandt, D.R., Miller, G.R., & Hocking, J.E. (1980a). Effects of self-monitoring and familiarity on deception detection. *Communication Quarterly, 28,* 3–10.

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

Brandt, D.R., Miller, G.R., & Hocking, J.E. (1982). Familiarity and lie detection: A replication and extension. *Western Journal of Speech Communication, 46*, 276–290.

Buller, D.B., & Burgoon, J.K. (1996). Another look at information management: A rejoinder to McCornack, Levine, Morrison, and Lapinski. *Communication Monographs, 63*, 92–97.

Buller, D.B., Comstock, J., Aune, R.K., & Strzyzewski, K.D. (1889). The effects of probing on deceivers and truthtellers. *Journal of Nonverbal Behavior, 13*, 155–170.

Buller, D.B., Stiff, J.B., & Burgoon, J.K. (1986). Behavioral adaptation in deceptive transactions: Fact or fiction: Reply to Levine and McCornack. *Human Communication Research, 22*, 589–603.

Buller, D.B., Strzyzewski, K.D., & Comstock, J. (1991a). Interpersonal deception I: Deceivers' reactions to receivers' suspicion and probing. *Communication Monographs, 58*, 1–24.

Buller, D.B., Strzyzewski, K.D., & Hunsaker, F.G. (1991b). Interpersonal deception II: The inferiority of conversational participants as deception detectors. *Communication Monographs, 58*, 1–24.

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

Burgoon, J.K., & Buller, D.B. (1996b). Reflections on the nature of theory building and the theoretical status of interpersonal deception theory. *Communication Theory, 6*, 310–328.

Burgoon, J.K., Buller, D.B., Dillman, L., & Walther, J.B. (1995). Interpersonal deception IV: Effects of suspicion on perceived communication and nonverbal behavior dynamics. *Human Communication Research, 22*, 163–196.

Burgoon, J.K., Buller, D.B., Ebesu, A.S., & Rockwell, P. (1994). Interpersonal deception V: Accuracy in deception detection. *Communication Monographs, 61*, 303–325.

Burgoon, J.K., Buller, D.B., Ebesu, A.S., White, C.H., & Rockwell, P.A. (1996). Testing interpersonal deception theory: Effects of suspicion on communication behaviors and perceptions. *Communication Theory, 6*, 243–267.

Burgoon, J.K., Buller, D.B., Guerrero, L.K., Afifi, W.A., & Feldman, C.M. (1996). Interpersonal deception XII: Information management dimensions underlying deceptive and truthful messages. *Communication Monographs, 63*, 50–69.

Burgoon, J.D., Buller, D.B., & Woodall, W.G. (1989). *Nonverbal Communication: The Unspoken Dialogue.* NY: Harper & Row.

Christie, R., & Geis, F.L. (1970). *Studies in Machiavellianism.* NY: Academic Press.

Comadena, M.E. (1982). Accuracy in detecting deception: Intimate and friendship relationships. In M. Burgoon (Ed.) *Communication Yearbook 6* (pp. 446–472). Beverly Hills, CA: Sage.

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

DePaulo, B.M., Kirkendol, S.E., Tang, J., & O'Briend, T.P. (1988). The motivation impairment effect in the communication of deception: Replications and extensions. *Journal of Nonverbal Behavior*, 12, 177–202.

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.

deTurck, M.A., & Miller, G.R. (1990). Training observers to detect deception: Effects of self-monitoring and rehearsal. *Human Communication Research, 16*, 603–620.

Ekman, P., & Friesen, W.V. (1974). Detecting deception from the body or face. *Journal of Personality and Social Psychology, 20*, 288–298.

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

Fay, P.J., & Middleton, W.C. (1941). The ability to judge truth-telling, or lying from the voice transmitted over a public address system. *The Journal of General Psychology, 24*, 211–215.

Feeley, T.H., deTurck, M.A., & Young, M.J. (1995). Baseline familiarity in lie detection. *Communication Research Reports, 12*, 160–169.

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

Geizer, R.S., Rarick, D.L., & Soldow, G.F. (1977). Deception and judgment accuracy: A study in person perception. *Personality and Social Psychology Bulletin, 3*, 446–449.

Gilbert, D.T., Krull, D.S., & Malone, P.S. (1990). Unbelieving the unbelievable: Some problems in the rejection of false information. *Journal of Personality and Social Psychology, 59*, 601–613.

Kalbfleisch, P.J. (1994). The language of detecting deceit. *Journal of Language and Social Psychology, 13*, 469–496.

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

Kraut, R., & 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.

Levine, T.R., & McCornack, S.A. (1992). Linking love and lies: A formal test of McCornack and Parks' model of deception detection. *Journal of Social and Personal Relationships, 9*, 143–154.

Levine, T.R., & McCornack, S.A. (1994). *Behavioral adaptation and heuristic message processing explanations of the probing effect.* Presented at the annual meeting of the International Communication Association, Sydney, Australia.

McCornack, S.A. (1992). Information manipulation theory. *Communication Monographs, 59*, 1–16.

McCornack, S.A. (1997). The generation of deceptive messages: Laying the groundwork for a viable theory of interpersonal deception. In J.O. Greene (Ed.). *Messages Production.* Mahwah, NJ: LEA.

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.* Beverly Hills, CA: Sage.

McCornack, S.A., & Levine, T.R. (1990). When lovers become leery: The relationship between suspicion and accuracy in detecting deception. *Communication Monographs, 57*, 219–230.
Millar, M., & Millar, K. (1995). Detection of deception in familiar and unfamiliar persons: The effects of information restriction. *Journal of Nonverbal Behavior, 19*, 69–84.
Millar, M.G., & Millar, K.U. (1997). The effects of cognitive capacity and suspicion on truth bias. *Communication Research, 24*, 556–570.
Miller, G.R., & Stiff, J.B. (1993). *Deceptive Communication*. Newbury Park, CA: Sage.
O'Sullivan, M., Ekman, P., & Friesen, W.V. (1988). The effect of comparisons on detecting deceit. *Journal of Nonverbal Behavior, 12*, 203–215.
Stiff, J.B., Kim, H.J., & Ramesh, C. (1992). *Truth biases and aroused suspicion in relational deception. Communication Research, 19*, 326–345.
Stiff, J.B., & Miller, G.R. (1986). "Come to think of it . . . ": Interrogative probes, deceptive communication, and deception detection. *Human Communication Research, 12*, 339–357.
Stiff, J.B., Miller, G.R., Sleight, C., Mongeau, P.A., Garlick, R., & Rogan, R. (1989). Explanations for visual cue primacy in judgments of honesty and deceit. *Journal of Personality and Social Psychology, 56*, 555–564.
Toris, C., & DePaulo, B.M. (1985). Effects of actual deception and suspiciousness of deception on interpersonal perceptions. *Journal of Personality and Social Psychology, 47*, 1063–1073.
Turner, R.E. Edgley, C., & Olmsteead, G. (1975). Information control in conversations: Honesty is not always the best policy. *Kansas Journal of Sociology, 11*, 69–89.
Vrij, A. (1994). The impact of information and setting on detection of deception by police detectives. *Journal of Nonverbal Behavior, 18*, 117–136.
Zuckerman, M., DePaulo, B.M., & Rosenthal, R. (1981). Verbal and nonverbal communication of deception. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 14, pp. 1–59). New York: Academic Press.
Zuckerman, M., Koestner, R., Colella, M.J., & Alton, A.O. (1984). Anchoring in the detection of deception and leakage. *Journal of Personality and Social Psychology, 47*, 301–311.