Communication Monographs
Vol. 73, No. 3, September 2006, pp. 243–260

Routledge
Taylor & Francis Group

Deception Detection Accuracy is a Predictable Linear Function of Message Veracity Base-Rate: A Formal Test of Park and Levine's Probability Model

Timothy R. Levine, Rachel K. Kim, Hee Sun Park & Mikayla Hughes

This study provided the first empirical test of point predictions made by the Park-Levine probability model of deception detection accuracy. Participants viewed a series of interviews containing truthful answers, unsanctioned, high-stakes lies, or some combination of both. One randomly selected set of participants (n = 50) *made judgments where the probability that each message was honest was P(H) = .50. Accuracy judgments in this condition were used to generate point predictions generated from the model and tested against the results from a second set of data* (n = 413)*. Participants were randomly assigned to one of eight base-rate conditions where the probability that a message was honest systematically varied from 0.00 to 1.00. Consistent with the veracity effect, participants in P(H) = .50 condition were significantly more likely to judge messages as truths than as lies, and consequently truths (67%) were identified with greater accuracy than lies (34%). As predicted by the model, overall accuracy was a linear function of message veracity base-rate, the base-rate induction explained 24% of the variance in accuracy scores, and, on average, raw accuracy scores for specific conditions were predicted to within approximately ± 2.6%. The findings show that specific deception detection accuracy scores can be precisely predicted with the Park-Levine model.*

Keywords: Deception; Accuracy; Lying; Lies; Truth-bias

Timothy R. Levine (PhD, Michigan State University, 1992) is Professor in the Department of Communication at Michigan State University where Rachel K. Kim (MA, University of Hawaii, 2004) is a doctoral student and Hee Sun Park (PhD, University of California, Santa Barbara, 2003) is Assistant Professor. Mikayla Hughes (PhD, Michigan State University, 2006) is Assistant Professor of Communication at University of California, Davis. Correspondence to Tim Levine, Department of Communication, Michigan State University, East Lansing, MI, USA. Tel: 48824-1212; Email: levinet@msu.edu

ISSN 0363-7751 (print)/ISSN 1479-5787 (online)
DOI: 10.1080/03637750600873736

Much research exists assessing people's (in)ability to accurately distinguish truths from lies, and the results of these investigations are highly consistent from study to study. Deception detection accuracy is significantly, but only slightly, better than 50-50. The across-study accuracy average is 54% (Bond & DePaulo, 2006). This conclusion is both widely accepted (Levine, Park, & McCornack, 1999) and very well documented (Bond & DePaulo). It is also potentially misleading because this conclusion is likely limited to specific research design idiosyncrasies that transcend the literature, and may not generalize to nonresearch settings.

Park and Levine (2001) developed a probability model of accuracy in deception detection experiments that makes specific and testable predictions about how accuracy varies as a function of message veracity base-rate. Consistent with Levine et al.'s (1999) veracity effect findings, the model predicts a linear relationship. To the current authors' knowledge, however, the model's predictions have not been put to formal, rigorous, empirical test. If the model is valid, the model provides a high degree of predictive precision because it can be used to predict specific accuracy levels under specifiable conditions.

The goal of this paper is to expose the Park-Levine model to a series of increasingly stringent tests. First, a set of high-quality stimulus materials was created. The lies evaluated were high-stakes and unsanctioned, yet ground truth was fully known. Next, a randomly selected group of participants made veracity judgments about a series of messages where half were true and half were lies. These data were used to test veracity effect predictions and to generate point estimates of accuracy rates under different base-rates. These predictions were then tested on data collected from a different group of participants who were randomly assigned to one of eight different base-rate conditions. A comparison of the accuracy levels observed in each base-rate condition to those predicted by the model provided a stringent test of the predictive power of the Park and Levine (2001) model.

Deception Detection Accuracy

More than 200 studies have investigated people's ability to accurately detect deception, and most of these experiments share common design features. Participants are typically shown a series of messages and asked to judge veracity. Repeated measures are most common, but sometimes independent groups designs are used. Participants are most often asked to make dichotomous truth–lie judgments, but some studies scale deception judgments. Meta-analysis, however, shows that dichotomous and continuous scaling of lie judgments yield comparable results (Bond & DePaulo, 2006).

The primary conclusion drawn from the literature is that people are not especially accurate in detecting deception. The most recent and thorough meta-analysis finds an across-study average of 54% accuracy (Bond & DePaulo, 2006). This rate is significantly above 50%, but unimpressive on the face. Known moderators exist, but they typically make only a small difference. Instead, research findings are very consistent in an absolute sense, the standard errors in accuracy judgments tend to be small, and presumably more reliable studies involving larger numbers of judgments tend to report findings closer to the across-study average (Bond & DePaulo).

A second reliable finding in the literature is that people are typically truth-biased (Levine et al., 1999). A truth-bias is a tendency to judge messages as truthful rather than deceptive independent of actual message veracity. It is possible that a lie-bias might exist in certain populations or situations (Meissner & Kassin, 2002), but truth-bias is a much more common finding. Bond and DePaulo's (2006) meta-analysis finds that 56% of messages were judged as honest, but there is greater study-to-study variability in the percentage of truth judgments than in accuracy rates. Explanations for truth-bias include the cognitive processes involved in mentally representing true and false information (Gilbert, 1991) and fundamental principles in how language is understood (Grice, 1989; McCornack, 1992). Truth-bias is especially pronounced when people interact face-to-face (Buller, Strzyzewski, & Hunsaker, 1991), when people know the message source (McCornack & Parks, 1986), and in studies where the participants are unaware that their task is to detect lies (e.g., Levine et al., 2000; McCornack & Levine, 1990).

Veracity Effect Predictions

Levine et al. (1999) observed that the research documenting the accuracy-is-slightly-above-50% conclusion typically share two important features. The messages serving as stimulus materials in the research almost always have a 50-50 chance of being honest or deceptive, and detection accuracy is calculated by averaging across both truths and lies. These features may limit the generality of findings in important ways. Because people typically make truth judgments with greater frequency than lie judgments, they are more likely to be correct when judging truths as opposed to lies. Levine et al. label this robust phenomenon "the veracity effect," and an implication is that the findings observed in the literature are likely limited to the common .50 truth-lie base-rate that is nearly universal in the literature. That is, average accuracy rates in studies or situations where messages have a 50-50 probability of being honest or deceptive most likely do not generalize to situations where a markedly different base-rate exists. Consistent with this, Levine et al. reported accuracy of approximately 60% when 75% of messages were honest, 52% when 50% of messages were honest, and 40% when 25% of the messages judged were honest. Importantly, the observed relationship between message base-rate and accuracy was almost perfectly linear.

The veracity effect refers to the finding that detection accuracy is function of message veracity, and it has been suggested that message veracity may be the single strongest known predictor of detection accuracy (Levine et al., 1999). The reasoning and evidence behind the veracity effect rest on the presumption that people are most often truth-biased for the reasons noted above. If people are more likely to judge messages as truths than lies, then they are more likely to be accurate when judging truthful messages than lies. Levine et al. found evidence consistent with, and the current research is expected to replicate, the following hypotheses.

H1 Under conditions of an equal number of truths and lies, people will judge a greater proportion of the messages as honest than dishonest.

Consequently,

H2 Truth accuracy will be greater than lie accuracy.

Further,

H3 Truth accuracy will be significantly greater than 50% (H3a) and lie accuracy will be significantly below 50% (H3b).

If the above hypotheses are correct, then it follows that as the ratio of honest to deceptive messages under scrutiny should affect detection accuracy such that a greater proportion of honest messages should be associated with higher detection accuracy while a relatively larger proportion of lies would be associated with lower accuracy. Thus, an implication of the veracity effect is that base-rate impacts detection accuracy.

H4 *The truth-lie base-rate will affect detection accuracy such that the proportion of messages that are true will be positively associated with detection accuracy.*

The Park and Levine Probability Model

As an extension of the work on the veracity effect, Park and Levine (2001) modeled detection accuracy as a function of truth accuracy, lie accuracy, and the message veracity base-rate. The model views messages and veracity judgments as either truths or lies.[1] An accurate judgment is one in which either a truthful message is judged as truthful or a deceptive message is judged as deceptive. Otherwise, the judgment is inaccurate. Further, the model assumes that within a particular setting or context, the probability that a given message is judged as truthful is independent of the probability that the message is truthful. Truth-bias, however, is expected to vary across situations.

Let the set of truthful messages be T, lies $\sim$T (not truthful), messages judged as honest H, and messages judged as lies $\sim$H (not honest). Accurate judgments are reflected by the intersection of sets $T \cap H$ or $\sim T \cap \sim H$. Thus, accuracy is the union of honest messages judged as honest and lies judged as lies, i.e., $(T \cap H) \cup (\sim T \cap \sim H)$. Further, the probability that a message is truthful is P(T) and the probability that a message is judged as honest is P(H). P(T) is what has been referred to in this paper as base-rate and P(H) is truth-bias. The probability of a correct judgment is $P(T \cap H) + P(\sim T \cap \sim H)$.

The probability of the intersection of sets $T \cap H$ is equal to the conditional probability of a truth judgment given a true message times the probability that a message is true. This can be written as $P(T \cap H) = P(H|T) \times P(T)$. Similarly the probability of correctly identifying a lie $P(\sim T \cap \sim H)$ is equal to the conditional probability of a lie judgment given a deceptive message $P(\sim H| \sim T)$ times the probability that the message is a lie $P(\sim T)$. Total accuracy is simply a sum of these two probabilities: $P(H|T) \times P(T) + P(\sim H| \sim T) \times P(\sim T)$. Thus, the Park-Levine model predicts that the observed total accuracy will be the product of truth accuracy times the proportion of messages that are true plus the product of lie accuracy times the proportion of messages that are lies where the proportion of true messages equals one minus the proportion of messages that are lies. Park and Levine (2001) show this model predicts a linear relationship between the truth-lie base-rate and detection accuracy, and the relationship will be positive when truth accuracy exceeds lie accuracy. Thus, if H1 and H2 are consistent with the data, then the following is predicted.

H5 The relationship between the ratio of truthful to total messages and detection accuracy is linear and positive.

Further, the model can be tested by generating truth and lie accuracy values from one set of data, assessing the accuracy of a different set of participants exposed to varying base-rates, and comparing predicted accuracy rates to those obtained in the second set of data. That is, values for $P(H|T)$ and $P(\sim H|\sim T)$ can be estimated from one set of participants, and values of $P(T)$ and $P(\sim T)$ can be experimentally varied. Knowing these values, point estimates of raw accuracy rates under different base rates can be generated and compared to actual data. To the extent to which the model has empirical value, the observed values should approximate those predicted by the model. The extent to which the predicted values match the observed values might be assessed in a variety of ways. These include, in order of increasingly stringent criteria, the following.

H6a There will be a strong positive correlation between predicted and obtained values.

H6b The slope and the y-intercept of the straight line generated from the predicted values will fall within the 95% confidence intervals of the ordinary least squares regression line generated from the data.

H6c The observed accuracy values in each base-rate condition will not significantly differ from the value predicted by the model.[2]

Obviously, due to sampling error and other contaminants, even if the model is accurate, observed values will depart from those predicted by the model to some extent. A research question is therefore proposed to ascertain, in descriptive terms, what degree of precision is observed in the model test.

RQ1 In terms of raw percentages, how closely can detection accuracy rates be predicted from truth-lie base rates?

Method

Overview

This research was done in several phases. First, a number of truthful messages and lies were videotaped for use in the deception detection portion of the data collection. The videotapes were digitized, segmented, and formatted into sequences with different base-rates. A different set of participants viewed the videotaped messages and made veracity judgments, which were scored for accuracy. The participants making the veracity judgments were randomly assigned to either a 50-50 truth–lie control condition, or one of eight different base-rate conditions ranging from zero to 100% honest. Hypotheses 1–3 were tested on the data from the control condition and Hypotheses 4 and 5 were tested on the data involving the base-rate induction. Model fit (H6) was tested by using the data in the control condition to generate point predictions that were tested against the results obtained in the base-rate induction data. All phases of data collection were IRB approved.

Stimulus Material Creation

The first phase involved the creation of a series of truthful and deceptive messages for use in subsequent hypothesis and model tests. In obtaining these messages, several criteria were considered. First, ground truth needed to be known with absolute certainty and the veracity of message content needed to be unambiguously truthful or deceptive. That is, in order to meaningfully calculate accuracy, it must be assured that the truthful messages were objectively true and lacking in deceptive intent, and that the lies were in fact lies (false statements designed to mislead). Second, unsanctioned lies were preferred. Sanctioned lies are made in response to researcher instruction whereas unsanctioned lies are ones in which the message sources decide to lie or not. Although most previous deception research has involved sanctioned lies, unsanctioned lies are desirable for reasons of ecological validity and diagnostic utility (Miller & Stiff, 1993). Relative to unsanctioned lies, sanctioned lies may be associated with lower levels of guilt and arousal making them potentially less different from truthful messages. For similar reasons, relatively high-stakes consequences were required. High-stakes lies are presumed to be more arousing, and nonverbal clues to deceit are more likely to be apparent in high-stakes situations (DePaulo et al., 2003; Frank & Feeley, 2003). If lies were sanctioned or inconsequential, then low accuracy might be attributable to a theoretical lack of difference between truths and lies. Other validity-relevant criteria considered included audio and video playback quality and the availability of a sufficient amount of content upon which to make a veracity judgment. In order to meet these criteria, a variation on the Exline procedure (Exline, Thibaut, Hickey, & Gumpert, 1970) was employed.

Sixty-eight undergraduates participated in the message generation task, although the first eight sessions were run as practice and were not used to create the stimulus materials. The participants were recruited from a large basic course that enrolls largely freshman nonmajors. The study was referred to as the "trivia game study," and participants were told that the purpose of the study involved investigating teamwork processes. Each experimental session involved four individuals: the actual participant, hereafter P, the confederate, C, the experimenter, E, and the principal investigator, PI. The roles of C, E, and PI were played by the same individuals throughout, and the behaviors of each were scripted, well rehearsed, and held constant.

Ps arrived at the lab individually and were paired with a research confederate who they believed to be another subject and their partner in the experiment. The same female C was used throughout, and none of the Ps reported suspecting that the confederate was anything other than another participant.[3] Ps were greeted by the male PI, and were introduced to a female E who gave instructions, administered the trivia game, and conducted a postgame videotaped interview. Ps were seated at a small table next to C, across from E, and with their back to the door. The PI left the room once the consent procedure and the instructions began.

All participants played a trivia game for a monetary prize in addition to standard research credit. They were told that they would be working as a team with another participant, and that the team who answered the most questions correctly would win $20 each.[4] Four different sets of trivia questions were used, but the questions were uniformly very difficult. Few participants knew the answers to more than one of the 10 questions.

Between the third and fourth questions, a cell phone ring could be heard in an adjoining room, followed by the muffled voice of the PI. The PI then burst into the room where the trivia game was in progress, and told E that there was an emergency phone call from daycare, that the call was in reference to the E's son, and that E needed to take the call immediately. The PI told P and C to wait in the room, and the PI and E rushed out, loudly closing a series of three doors behind them. The answers to the trivia questions were left in a folder on the desk where the experimenter had been sitting. It was at this point that the cheating induction took place.

According to a randomized, counterbalanced, and predetermined schedule, the C attempted to instigate cheating during more than half of the sessions.[5] In the cheating conditions, C noted that she believed that the answers were in the folder on the desk, that she desired the monetary reward, and proposed that she and P cheat in order to improve their scores and win the money. C did not excessively pressure reluctant Ps. In the no cheating condition, the C did not attempt to instigate cheating, and engaged in small talk with P if P initiated talk. Otherwise, C studied. Both E and PI were blind to condition.

Approximately half of the Ps in the cheating condition actively participated in cheating.[6] Of those, approximately half denied cheating during the interview while the rest confessed under interrogation. Ps in the cheating condition who did not actively cheat or who later confessed were not used in the present investigation. No P in the honest condition attempted to cheat, and thus all denials in the noncheating condition were objectively honest.

After about 5 minutes, E and PI returned, and the return was foreshadowed to Ps by the sound of doors being opened. After an apology and explanation, the trivia game resumed. Following the last question, E informed P and C that they would be interviewed separately, with E interviewing P and PI interviewing C in an adjoining room. P was seated in a chair and given a lapel microphone. A video camera on a tripod was positioned across the room. The camera was directly in front of P and a full body view was captured. The angle and zoom were held constant across interviews. Once the camera was turned on, E sat at a 30 degree angle to P just out of the picture, and asked a series of preplanned questions.[7] Following the interview, the PI debriefed P, and obtained permission to use the taped interview in subsequent research.[8] The PI probed if Ps were suspicious and the PI swore Ps to secrecy.

The videotapes were digitized and segmented into interviews. Fifteen usable honest interviews and seven usable deceptive interviews were obtained. Confessions and veracity-ambiguous statements were not used (see note 6). Once the tapes were digitized and segmented, nine different sets of interviews were created and burned onto DVDs for use in the second phase. Each interview lasted approximately 2 minutes.

First, a 50-50 truth–lie DVD was created. This video contained all seven lies and seven randomly selected truthful messages, and thus contained 14 messages total. The order of the clips was determined at random.

In order to vary message veracity base-rate, a DVD containing just the lies was created. The order of the interviews was determined at random. To create a second version with a different base-rate, one lie, at random, was omitted and replaced with a randomly selected truthful interview from the 50-50 compilation. This procedure was repeated without replacement until eight versions were created, each containing seven messages, and ranging from 0 to 7 lies.

Main Experiment

Participants. Four-hundred and sixty-three undergraduate students (50.8% female) participated in the deception detection experiment. Four-hundred and thirteen students participated in one of the base-rate conditions and fifty were in the offset control group. Most (80%+) participants were between 18 and 20 years old (mode = 18). All received class research credit in exchange for their participation.

Design. The design was a one-way, randomized, independent groups experiment with eight base-rate conditions plus an offset 50-50 base-rate control. Message veracity base-rate was the independent variable, and had eight levels ranging from 0 to 7 honest messages (i.e., 0.00, 0.14, 0.29, 0.43, 0.57, 0.71, 0.86, 1.00) with equal differences between adjacent levels. The offset control group had an equal number of truths and lies. Participants in the eight base-rate conditions watched and rated seven interviews while those in the control condition watched and rated fourteen interviews. Detection accuracy was the primary dependent variable.

Procedures and measures. Assignment to experimental condition was determined according to a preplanned, randomized, counterbalancing schedule. Each experimental session was randomly determined to be either one of the eight base-rate conditions or the control group. Participants signed up online for an experimental session, and sessions were assigned to condition based on the schedule. Up to 10 individuals could participate in a given session. The randomization schedule was followed until approximately 50 individuals had participated in a given condition. At that point, no additional data were collected in that condition and subsequent sessions were assigned to the next open condition on the schedule. Cell sizes ranged from $n = 49$ to 56.

Upon arrival to the laboratory, participants were seated at one of several desks in front of a large screen television. Participants were told the study was about "perceptions of others' communication." Instructions followed the consent procedure. All participants were given identical instructions which included the following:

> We are interested in people's perceptions of others' communication. You will be shown a series of interviews about the role of teamwork in a trivia game. The people on the videotape played a trivia game with a partner for a cash prize. During the game, the person administering the trivia questions unexpectedly had to leave the room. As a result, the people you will see were left alone in the room with the questions and answers to the trivia game along with their partners. After they completed the game, they were interviewed about their performance. The questions asked were always the same. They were asked to explain their performance and asked about the role of teamwork. They were also asked if they cheated in the game. We would like to know if you believe them about whether or not they cheated. After watching each segment, check whether or not you believe the person who was interviewed.

The questionnaire provided a forced-choice pair of response options for each segment. These read "I think the person was honest and did not cheat" and "I think the person cheated even though they deny it." Based on these responses, accuracy scores were created as the percentage or proportion correct.

Results

The first three hypotheses were tested on the data ($n = 50$) obtained from the 50-50 truth–lie control condition. The first hypothesis predicted that under conditions of an equal number of truths and lies, people would judge more messages as truths than as lies. The data were consistent with this prediction. Participants judged between 43% and 93% of the messages as true ($M = 66.1\%$, $SD = 14.4\%$). A one-sample t-test showed that the average number of truth judgments exceeded 50%, $t(49) = 7.92$, $p < .001$.

Total accuracy (scored across truths and lies) ranged from 29% to 71% ($M = 50.7\%$, $SD = 10.1\%$, 95% CI = 47.9–53.5%). A one-sample t-test showed that average accuracy did not exceed 50%, $t(49) = 0.50$, $p = .62$.

Hypothesis 2 posited that truth accuracy would be greater than lie accuracy. Truth accuracy scores ranged from 29% to 100% with a mean of 67.1% ($SD = 17.9\%$) and lie accuracy ranged from 0% to 71% ($M = 34.3\%$, $SD = 17.3\%$). A paired t-test showed that these means were significantly different, $t(49) = 8.08$, $p < .001$.

Hypothesis 3 predicted that truth accuracy would be significantly greater than 50% (H3a) and lie accuracy would be significantly below 50% (H3b). The data were consistent with both parts of the hypothesis. One-sample t-tests showed that truth accuracy was significantly greater than 50%, $t(49) = 6.79$, $p < .001$ and lie accuracy was significantly below 50%, $t(49) = -6.42$, $p < .001$.

Hypothesis 4 predicted that the truth–lie base-rate would affect detection accuracy such that the proportion of messages that are true will be positively associated with detection accuracy. This hypothesis was tested on the base-rate varied data ($n = 413$) with a one-way ANOVA. As anticipated, the base-rate induction had a substantial impact on accuracy ratings, $F(7, 405) = 18.08$, $p < .001$, $\eta^2 = .24$, $\eta = .49$. As shown in Table 1, accuracy means tend to increase as the proportion of messages that are true increases.

Table 1 Predicted and Observed Accuracy Rates by Message Veracity Base-Rate Condition

Base-Rate	Predicted	Actual	Deviation	t	95% CI
00%	.3429	.3641	−.0212	−0.74	.3066 to .4217
14%	.3898	.3557	.0341	1.37	.3060 to .4055
29%	.4368	.3915	.0453	1.73	.3390 to .4441
43%	.4837	.4752	.0085	0.29	.4158 to .5347
57%	.5306	.5602	−.0296	−1.06	.5041 to .6164
71%	.5775	.5663	.0112	0.45	.5169 to .6158
86%	.6245	.5882	.0363	1.26	.5305 to .6460
100%	.6714	.6514	.0200	0.92	.6079 to .6950

Note: The base-rate is the percentage of messages that were honest. None of the t-values are statistically significant at $p < .05$.

Hypothesis 5 further specified that the relationship between the ratio of truthful to total messages and detection accuracy is linear and positive. The data were consistent with this hypothesis. A linear contrast was statistically significant and substantial, $F(1, 405) = 119.56$, $p < .001$, $\eta^2 = .22$, $r = .47$, and accounted for 94.5% of the total explained sums of squares. Although one cell mean was out of order, the deviation from linearity was not statistically significant, $F(6, 405) = 1.16$, $p = .33$, $\eta^2 = .01$. The relationship was significantly positive, slope $= +0.316$, standardized regression coefficient, $\beta = +.474$, $p < .001$.

To assess model fit, predicted accuracy values were calculated for each base-rate level on the basis of the truth and lie accuracy results obtained in the control condition and the experimentally varied base-rate in each condition. The formula was .6714 times the number of messages that were true divided by 7 plus .3429 times the number of messages that were lies divided by 7. The predicted and obtained values are presented in Table 1 and Figure 1.

As visually depicted in Figure 1, the data closely approximated the values predicted by the model. Statistically, the degree of fit can be shown in a number of ways. Hypothesis 6a, for example, predicted a strong positive correlation between predicted and obtained values. This was obtained, $r(6) = .97$, $p < .01$.

Hypothesis 6b specified that the slope and the y-intercept of the straight line generated from the predicted values will fall within the 95% confidence intervals of the ordinary least squares (OLS) regression line generated from the data. The line produced from the predicted values was:

$$\text{accuracy} = .343 + .328\ (\text{base-rate})$$

The OLS line fitting the actual data was:

$$\text{accuracy} = .336 + .316\ (\text{base-rate}) + \text{error}$$

The 95% confidence interval around the observed intercept is .303–.369 and the interval around the slope is .259–.373. Thus, consistent with H6b, the predicted values fall within the confidence intervals around the obtained values. As an alternative test of H6b, confidence intervals were created around each cell mean. The

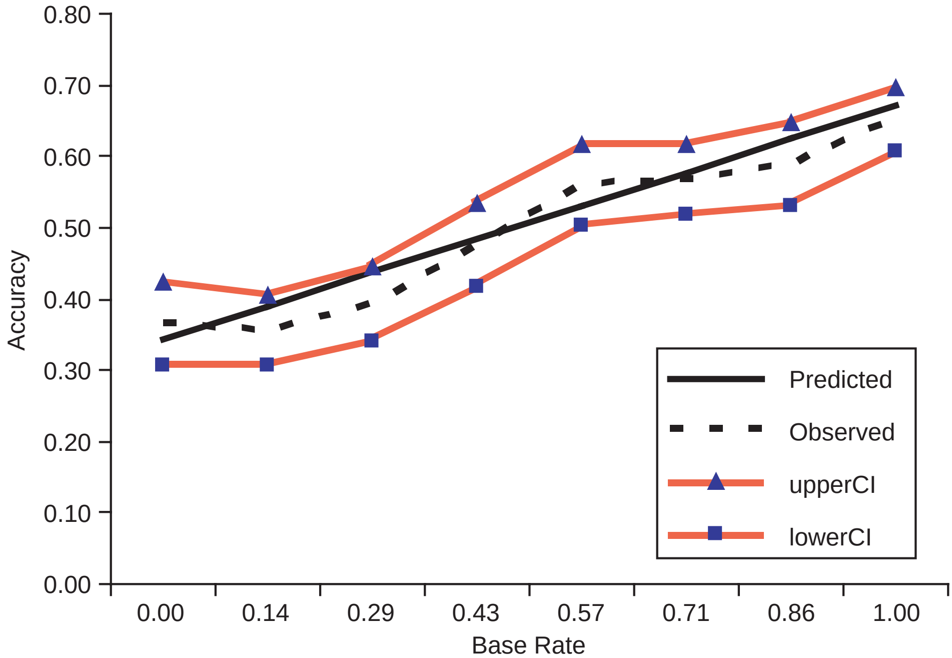

Figure 1 Observed and predicted accuracy rates as a function of message veracity base-rate.

predicted line, and the obtained means and confidence limits are plotted in Figure 1. Visual inspection shows the predicted values falling within the 95% confidence intervals of the obtained values.

It was further predicted in Hypothesis 6c that the observed accuracy values in each base-rate condition would not significantly differ from the value predicted by the model. This was tested with a series of one-sample t-tests. As shown in Table 1, none of the eight t-tests were statistically significant indicating that the obtained cell means in each of the eight cells did not differ significantly from the point predictions. The average absolute value of the difference between predicted and obtained was 0.0258. Thus, in answer to the research question, on average raw detection accuracy scores were predicted to within 2.6%. The largest deviation was 4.5% in the 29% honest condition.

Recall that an assumption of the model is that truth-bias is independent of base-rate. Examination of a one-way ANOVA shows that truth-bias did not significantly vary across base-rate conditions, $F(7, 405) = 0.86$, $p = .54$, $\eta^2 = .01$.[9]

Discussion

This paper sought to replicate Levine et al.'s (1999) veracity effect findings and provide a formal test of the Park and Levine (2001) model of deception detection accuracy. This was done by creating a series of videotaped truths and lies, experimentally varying the proportion of messages that were truthful, and having judges make honesty judgments. Accuracy was calculated, and obtained accuracy rates were shown to closely approximate those predicted by the model.

The veracity effect refers to the finding that the a priori probability of a message being honest is a strong predictor of detection accuracy (Levine et al., 1999). This happens because people tend to be truth-biased and consequently are more likely to correctly identify truths than lies (Levine et al.). Consistent with the veracity effect, participants in the control condition judged more messages as truths than as lies. Truth accuracy was greater than both lie accuracy and fifty percent while lie accuracy was below 50%. When the proportion of truthful messages was experimentally varied, accuracy increased as the proportion of honest messages increased, and the base-rate induction accounted for 24% of the variance in accuracy judgments. Thus, the current findings closely replicate the results of Levine et al. and allow for additional confidence in the veracity effect.

More importantly, however, this paper provided the first real test of Park and Levine's (2001) probability model. The model extends upon the veracity effect reasoning and depicts how variation in message veracity base-rate affects deception detection accuracy. The model specifies a linear relationship and makes precise predictions about how accuracy changes as a function of base-rate.

The current results closely approximated the predictions of the model. The observed relationship between base-rate and accuracy was linear. The line predicted by the model fell within the confidence intervals around the slope and intercept of the line reflecting the data. None of the observed cell means differed significantly from the predicted values, and observed and predicted values correlated at more than $r =$ .97. On average, cell means were predicted to within $\pm 2.6\%$, and all eight of the observed values were within 5% of the predicted values. Thus, the Park-Levine model appears to offer a level of predictive precision unusual in social science research.

The current findings, coupled with those of Levine et al. (1999) and Park and Levine (2001) extend the understanding of deception detection processes. Base-rate is a strong predictor of accuracy, accounting for 21% of the variance in Levine et al. and 24% of the variance in the current experiment. In terms of raw accuracy score, base-rate accounted for almost a 30% difference (from 36% accurate in the 0% honest condition to 65% accurate in the 100% honest condition). Few, if any, variables have such a strong and reliable effect on detection accuracy. Further, the current data show that deception detection accuracy is a predictable linear function of message veracity base-rate, and that the model can be used to make accurate point predictions of raw accuracy scores.

These base-rate effects are explainable by the existence of truth-bias which is in turn explainable in terms of how people mentally represent true and false information (Gilbert, 1991) and how people make sense of others' messages (McCornack, 1992). Theoretically, the effect of base-rate on accuracy would become stronger when communication is face-to-face, when interaction is with a close relational partner, and when people are not primed to be suspicious or prompted to expect deception (Buller et al., 1991; McCornack & Levine, 1990; McCornack &

Parks, 1986). Alternatively, previous research would suggest weaker results when judges receive detection training, when they have occupational experience in lie detection, or when they are incarcerated as these conditions lower truth-bias, but do not substantially increase overall accuracy (Bond, Malloy, Arias, Nunn, & Thompson, 2005; Frank & Feeley, 2003; Levine, Feeley, McCornack, Harms, & Hughes, 2005; Meissner & Kassin, 2002).

Truth-bias in the current study (66%) was substantially higher than the across-study average reported in meta-analysis (56%; Bond & DePaulo, 2006) but lower than that reported in previous veracity effect research (69–74%; Levine et al., 1999). Across-study variation in participant instructions is one likely explanation for variation in truth-bias levels. Explicitly instructing participants to look for lies likely lowers truth-bias. In the current research, participants were not specifically told upfront that the study involved deception detection.

Perhaps the most obvious implication of these findings are for the widely held and well documented belief that deception detection accuracy is slightly but significantly above 50% (cf. Bond & DePaulo, 2006; Miller & Stiff, 1993; Vrij, 2000). This conclusion is warranted and accurate when the probability that a message is honest is approximately 50-50. However, because almost all prior research fixes the message veracity base-rate at exactly .50, and because levels of raw accuracy vary predictably and substantially as a function of base-rate, prior findings are limited accordingly. So long as people are truth-biased, accuracy will be lower when messages are predominantly deceptive and may be much higher when most messages are honest. Further, the more truth-biased a set of judges, the stronger the base-rate effect.

So, just how general is the slightly above 50% accuracy conclusion in light of the current findings? We believe it is very general to the extant experimental research, but probably not very general to everyday lie situations. DePaulo, Kashy, Kirkendol, and Epstein (1996) report that people tell, on average, about one or two lies per day. When one considers the sheer number of messages people send and receive on a daily basis, it is possible and perhaps even likely that the everyday lie base-rate departs radically from 50-50. We speculate that most of the messages that most people send and receive are probably honest most of the time.

We would further argue that people are probably more truth-biased in everyday conversation than in the deception lab (cf. Farquhar, 2005). Research shows that people are more truth-biased when interacting face-to-face (Buller et al., 1991), when they know the person they are talking to (McCornack & Parks, 1986), and they are not forewarned of impending deception (e.g., Levine et al., 2000; McCornack & Levine, 1990). Each of these conditions is often met in everyday situations, but is seldom the case in deception detection experiments. If these effects are additive, people might often be exceptionally truth-biased in most nonresearch settings.

If this speculation is accurate, people may inadvertently be highly accurate in distinguishing truths from lies. The Park-Levine model would predict very high levels of overall accuracy under conditions of elevated truth-bias and predominantly honest communication. The occasional lie, however, is likely to slip past unnoticed, and most likely goes undetected (or perhaps is uncovered at a much later date, cf. Park, Levine, McCornack, Morrison, & Ferrara, 2002). Alternatively, if most messages are indeed deceptive, then accuracy rates may be very low. Whichever is the case, however, message veracity base-rates likely vary from situation to situation and probably seldom approximate 50% except in experimental settings.

Whereas the slightly better than 50-50 accuracy conclusion is likely limited to the laboratory, the more general conclusion that accuracy rates are just above chance is much more general, depending on what is meant by chance. If base-rates and degree of truth-bias are considered, then the accuracy of participants at all base-rates are close to chance. For example, across base-rate conditions, 66.7% of messages were judged as honest. On pure chance, approximately two-thirds of truths and one-third of lies would be correctly identified. These values closely approximate the results obtained in the $P(H) = 0.00$ and $P(H) = 1.00$ conditions. Thus, while the generality of the 50%+ conclusion is questioned, the just above chance conclusion is not necessarily questioned. In fact, the thesis of this work is that chance explains deception detection rates well.

Another important set of implications stem from the model's assumptions that truth-bias varies across situations, but that individuals are insensitive to differences in base-rates within fixed situations. As previously noted, research documents that truth-bias is higher in face-to-face interaction and with relationally close others, and lower when information from third parties prompts suspicion. Further, it is likely that individuals have schemas or stereotypes concerning certain situations in which one should be alert to the possibility of deception, and people are likely less truth-biased in situations where they perceive that another person might be motivated to deceive (cf. Hilton, Fein, & Miller, 1993). Thus, the truth and lie accuracy discrepancy should vary predictably across situations, and understanding situational determinants of truth-bias is a central concern in forecasting truth and lie detection accuracy.

Alternatively, the Park-Levine model predictions assume that truth-bias is unrelated to base-rate within situations. The implication here is that source verbal and nonverbal behaviors play relatively little role in predicting deception detection accuracy. Whether because source behaviors have little actual diagnostic utility, or because message judges fail to use behavioral observations in such a way that systematically impacts accuracy, or both, the current data show that the Park-Levine model can predict accuracy with considerable precision without consideration of behavioral deception cues. This contrasts dramatically with virtually all alternative theoretical accounts for detection accuracy which emphasize the role of source verbal and especially nonverbal behavior in detection accuracy.

Given that the current study was done in the laboratory and that judges did not interact with the message sources face-to-face, some might question if the current findings would generalize to interactive situations. Although this is an empirical question, good reasons exist to believe that the model would likely hold in interactive settings, and that the effects for base-rate would be even stronger. Research on interactive deception and the effects of questioning find little effects on accuracy but a general increase in truth-bias (e.g., Buller et al., 1991; Burgoon, Buller, & Floyd, 2001;

Levine & McCornack, 2001). Thus, the effects for base-rate would likely be more pronounced if this study were replicated with an interactive design.

It also might be questioned if the current results would hold for lie-biased judges. Although this too is an empirical question, it is anticipated that the model would hold, but the model would then predict a negative linear effect for base-rate on accuracy. More generally, the degree of truth- or lie-bias should be associated with a steeper slope obtained from the regression of base-rate onto accuracy with the sign of the slope being a function of the direction of the bias. If there were no bias in either direction, base-rate would be immaterial and accuracy levels at a 50-50 base-rate would generalize to other base-rates. Higher levels of accuracy at the 50% base-rate would translate into a greater y-intercept (assuming a fixed slope). Thus, the model should accommodate both truth- and lie-biases, and differential accuracy rates.

Two particular strengths of the research design and analysis merit comment. First, the quality of stimulus materials judged reflects an improvement over much of the literature. The use of the Exline procedure to generate unsanctioned and higher stakes lies allows for greater confidence in the validity of the results, as does the DVD quality video and the longer and more substantial answers.

Second, the current research involved making and testing point predictions in addition to simply using traditional null hypothesis significance tests. Making and testing point predictions is unusual in the softer social sciences, and involves riskier tests in that the predictions are more precise and the probability that a false model will be rejected is enhanced (cf. Meehl, 1967). That a model making point predictions was advanced, and that such a model was capable of generating reasonably accurate predictions is noteworthy.

A limitation in the study is possible selection effects in the stimulus materials related to not all message sources in the cheating condition lying. The authors believe that while the situation was suboptimal, selection effects are unlikely to have substantial impact on the results. Nevertheless, assignment to truth-lie condition was not random.

A second more serious limitation stems from the inferential difficulties involved in assessing model fit with significance testing. Data consistent with the model is indicated by nonsignificant differences, and inferring support for a non-nil null with traditional significance tests is problematic. A better test might involve equivalence testing with noncentral distributions.

In conclusion, this paper successfully replicated previous veracity findings and reported data consistent with the Park-Levine probability model of deception detection. Message judges were found to be truth-biased and consequently truthful messages were correctly identified as such more often than lies. Truth accuracy was significant above 50% and lie accuracy was significant below 50%. Message veracity base-rate had a substantial, positive, and linear effect on accuracy, observed accuracy levels were within sampling error of predicted values, and accuracy was predicted to within $\pm 5\%$ in all eight cells. These results suggest that accuracy at detecting deception depends on base-rate and commonly held conclusions about accuracy are often limited to situations where half the messages are true and half are lies.

Notes

[1] Readers might object to the model on the grounds that messages, judgments of messages, or both, might be thought of in terms of gradations of honesty rather than in discrete, dichotomous terms. Whereas it is the case that deception research sometimes involves scaling honesty-deception along a continuum, virtually all conceptual definitions of lies or deception (including the one adopted here) specify a dichotomous construct. A message is either intended to mislead another or it is not. Further, when judgments of deceit are scaled on a continuum, such judgments likely confound perceptions of deceptive intent, the proportion of truthful message content, confidence in judgment, and perceived lie severity (Levine, 2001; Levine, Asada, & Massi, 2003).

[2] Readers might object to H6c on the grounds that it specifies a null hypothesis, and that null hypotheses cannot be adequately assessed with traditional null hypothesis significance testing. Significance testing, however, is not being used in the usual way here. Whereas the null hypothesis is typically a nil null and reflects the negation of a researcher's prediction, here point predictions are being tested, and this constitutes a stronger use of significant testing than is typical (see Meehl, 1967, for an explanation). This is expected to be the most stringent test of the model because detection accuracy scores have very small standard errors (Bond & DePaulo, 2006). Thus, the model might fit very well and still fail this final test.

[3] One exception was a single student who had met the confederate in another research project where the confederate was the researcher. This student did not participate in the study.

[4] In actuality, \$20 was awarded to the best noncheater, and a second \$20 prize was awarded to a randomly selected cheater.

[5] As data collection progressed, it became apparent that many participants either were not cheating, or were confessing. Starting with the 49th session, all sessions were assigned to the cheating condition.

[6] The approximately 50% cheating rate differs dramatically from what is reported in the literature. Miller and Stiff (1993) report that approximately 90% of students in their studies cheat in the cheating condition. There are a number of possible reasons for the difference. The current students may have been less prone to cheating, or more suspicious of a set-up. Perhaps the monetary incentive was insufficient for some participants. Most likely, however, are two design decisions. First, for ethical reasons, the confederate did not strongly pressure reluctant participants. Second, a different definition of cheating may have been used. In the current study, only participants who actively cheated were retained. Specifically, a number of participants conspired with the confederate, but used the confederate to do the actual cheating. These participants were not retained for this study because the truth or falsity of their interview answers was ambiguous. For example, when asked to explain their success, some answered that C knew the answers (true). When asked if cheating occurred, they responded that they had not cheated (literally true). Interview answers that were technically true but functionally deceptive were not included in the present stimulus tapes. Most confessions occurred in response to the final question (see note 7).

[7] The interview questions included the following.

1. Did you find the trivia questions difficult?
2. Was teamwork much of a factor? How so? Please explain?
3. In looking at your score, you did better than other groups. Does this surprise you?
4. How would you explain your success?
5. Did cheating occur when I left the room?
6. Why should I believe you?

[8] In the debriefing, it was clear that the participants viewed the situation as one of consequence. Fear of detection was a frequent factor listed by noncheaters in the cheating

condition, those who confessed frequently expressed relief that no sanctions were forthcoming, and most liars maintained their innocence throughout the debriefing.

[9] Cell means and standard deviations were .64 (.20), .65 (.18), .65 (.18), .70 (.18), .66 (.14), .70 (.17), .68 (.22), and .65 (.15) for the eight base-rate conditions ranging from all lies to all truths. The control group values were $M = .66$, $SD = .14$.

References

Bond, C. F., Jr., & DePaulo, B. M. (2006). Accuracy of deception judgments. *Review of Personality and Social Psychology*, *10*, 214–234.

Bond, G. D., Malloy, D. M., Arias, E. A., Nunn, S. N., & Thompson, L. A. (2005). Lie-biased decision making in prison. *Communication Reports*, *18*, 1–11.

Buller, D. B., Strzyzewski, K. D., & Hunsaker, F. G. (1991). Interpersonal deception II: The inferiority of conversational participants as deception detectors. *Communication Monographs*, *58*, 25–40.

Burgoon, J. K., Buller, D. B., & Floyd, K. (2001). Does participation affect deception success? A test of the interactivity principle. *Human Communication Research*, *27*, 503–534.

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

DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbrick, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin*, *129*, 74–118.

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

Farquhar, M. (2005). *A treasury of deception*. New York: Penguin.

Frank, M. G., & Feeley, T. H. (2003). To catch a liar: Challenges for research in lie detection training. *Journal of Applied Communication Research*, *31*, 58–75.

Gilbert, D. T. (1991). How mental systems believe. *The American Psychologist*, *46*, 107–119.

Grice, P. (1989). *Studies in the way of words*. Cambridge, MA: Harvard University Press.

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

Levine, T. R. (2001). Dichotomous and continuous views of deception: A reexamination of deception ratings in information manipulation theory. *Communication Research Reports*, *18*, 230–240.

Levine, T. R., Anders, L. N., Banas, J., Baum, K. L., Endo, K., Hu, A. D. S., & Wong, N. C. H. (2000). Norms, expectations, and deception: A norm violation model of veracity judgments. *Communication Monographs*, *67*, 123–137.

Levine, T. R., Asada, K. J., & Massi, L. L. (2003). The relative impact of violation type and lie severity on judgments of message deceptiveness. *Communication Research Reports*, *20*, 208–218.

Levine, T. R., Feeley, T., McCornack, S. A., Harms, C., & Hughes, M. (2005). Testing the effects of nonverbal training on deception detection accuracy with the inclusion of a bogus train control group. *Western Journal of Communication*, *69*, 203–218.

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

Levine, T. R., Park, H. S., & McCornack, S. A. (1999). Accuracy in detecting truths and lies: Documenting the "veracity effect". *Communication Monographs*, *66*, 125–144.

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

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.

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, CA: Sage.

Meehl, P. E. (1967). Theory testing in psychology and physics: A methodological paradox. *Philosophy of Science*, *34*, 103–115.

Meissner, C. A., & Kassin, S. M. (2002). "He's guilty!": Investigator bias in judgments of truth and deception. *Law and Human Behavior*, *26*, 469–480.

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

Park, H. S., & Levine, T. R. (2001). A probability model of accuracy in deception detection experiments. *Communication Monographs*, *68*, 201–210.

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

Vrij, A. (2000). *Detecting lies and deceit*. New York: Wiley.