Journal of Applied Psychology
1987, Vol. 72, No. 4, 611–614

0021-9010/87/$00.75

Detecting Infrequent Deception

Kevin R. Murphy
Colorado State University

Recent proposals for using the polygraph and similar devices in routine screening have been aimed at detecting deception in situations sometimes characterized by low base rates. Equations are developed that show that extraordinarily high levels of accuracy would be needed to detect infrequent deception. In this context, the debate over the accuracy of these methods is irrelevant; the accuracy needed to detect infrequent deception far exceeds the levels claimed by the most enthusiastic proponents of these detection techniques. The limits on the use of any particular test of deception can be determined by considering the base rate for deception and the proportion of the nondeceptive population that fails the test. When the base rate is less than .10, these limits are extremely restrictive.

Methods of detecting deception, especially the polygraph, have been the subject of a great deal of research and debate, much of which has centered on the possibility that innocent parties will be wrongly labeled as deceptive (Lieblich, Ben-Shakhar, Kugelmass, & Cohen, 1978; Lykken, 1974, 1978, 1979, 1981, 1985; Podlensy & Raskin, 1977; Raskin & Podlensy, 1979; Sackett & Decker, 1979). Recently the American Psychological Association Council of Representatives passed a resolution on the polygraph that stated in part that

> There is the possibility of great damage to innocent persons who must inevitably be labeled as deceptors in situations where the base rate of deception is low; an unacceptable number of false positives would occur even should the validity of testing procedures be quite high. ("Council Takes Stand," 1986)

Polygraphs and similar methods are used in several contexts in which the base rate for deception is likely to be low. For example, nearly three fourths of all polygraph examinations are conducted for the purpose of preemployment screening (Kleinmuntz, 1985) and often focus on employee theft. The base rate for nontrivial employee theft (e.g., more than $5) is less than 5% in most settings (Hollinger & Clark, 1983), and the base rate for other types of nontrivial deception (except for questions involving drug use) in this context is thought to be equally low (U.S. Office of Technology Assessment, 1983).

It has long been known that base rates have a substantial impact on the accuracy of tests of every sort and that accurate detection of any condition (e.g., deception, psychopathology) for which the base rate is low is difficult (Dawes, 1962; Lykken, 1974; Meehl & Rosen, 1955; U.S. Office of Technology Assessment, 1983). The dual purpose of this article is to demonstrate the boundary conditions for detecting deception and to show that the accurate detection of infrequent deception demands impossibly high levels of validity.

I thank Jeanette Cleveland, George Thornton, and two anonymous reviewers for their valuable comments.

Correspondence concerning this article should be addressed to Kevin R. Murphy, Department of Psychology, Colorado State University, Fort Collins, Colorado 80523.

Formulation

The problem of detecting deception can be analyzed in terms of the base rate and the conditional probability of deception, given the results of a test of deception (Brett, Phillips, & Beary, 1986; Meehl & Rosen, 1955). Here, the base rate defines the prior probability of guilt, or deception. Thus, when 5% of all subjects are involved in illegal activities that they wish to conceal, the probability that a subject selected at random is so involved is .05. A polygraph, an honest test, a voice analysis, or some other method of detecting deception provides data that may lead to a revised estimate of the likelihood of deception. Finally, there must be some threshold for determining when the probability of deception is sufficiently high to conclude that the subject is, in fact, attempting to deceive the questioner.

One of the most useful equations for analyzing the problem at hand is

$$\frac{P(D|F)}{P(T|F)} = \frac{P(F|D)}{P(F|T)} \times \frac{P(D)}{P(T)}, \qquad (1)$$

where D is the hypothesis that the subject is deceptive, T indicates that the subject is telling the truth, F indicates that the subject fails the test and is labeled deceptive, $P(D)$ is the unconditional probability or base rate for deception, $P(D|F)$ is the conditional probability that the subject is deceptive, given the fact that the subject has failed the test, and $P(F|T)$ is the false positive rate (i.e., the proportion of nondeceptive subjects who failed the test).

Equation 1, which can be derived from the basic definition of conditional probability (c.f. Dawes, 1962), defines the odds form of Bayes Theorem. Equation 1 is equivalent to the statement *Posterior Odds = Likelihood Ratio × Prior Odds.* Here, the prior odds refer to the base rate, the likelihood ratio refers to the sensitivity of the test in discriminating deceptive from truthful subjects, and the posterior odds refer to the assessment, after the test, of the likelihood of deception. Prior and posterior odds are defined as the ratio of the probability that a person is deceptive to the probability that a person is not deceptive. When the odds are less than 1.0 (one to one), the evidence favors the hypothesis that the subject is not deceptive; odds greater than

Table 1
Maximum False Positive Rates That Would Allow Investigator To Meet Minimum Threshold for Deception

Base rate for deception	Maximum permissible false positive rate
.30	.428
.20	.250
.10	.111
.05	.052
.03	.030
.01	.010

1.0 indicate that the evidence favors the hypothesis that the subject is deceptive.

It is possible to use Equation 1 to define a minimum threshold for deciding whether a subject is deceptive, which, in turn, can be used to establish the boundary conditions for detecting infrequent deception. This formulation also provides simple methods of incorporating factors such as reasonable doubt and the possibility of false negatives in defining these boundary conditions.

Minimum Threshold

The minimum threshold for labeling a subject deceptive occurs when the data in favor of the hypothesis *deceptive* are at least as strong as the evidence against this hypothesis, which occurs when the posterior odds of deception are equal to 1.0. When the posterior odds are less than 1.0, the data favor the hypothesis *not deceptive;* it would be irrational to go against the weight of the available evidence and label a subject deceptive unless the potential damage associated with deception were very great (Ben-Shakhar, Lieblich, & Bar-Hillel, 1982; Lieblich, et al., 1978). As can be seen from Equation 1, posterior odds of 1.0 are attained when the likelihood ratio for a test is the reciprocal of the prior odds. For example, if the base rate for nontrivial employee theft were assumed to be .05, the prior odds that an employee chosen at random is a thief would be .052 (i.e., .05/.95), or 1 in 19. A test of deception would therefore have to have a likelihood ratio of 19 to provide any chance of concluding that this subject was deceptive. In concrete terms, this means that the true positive rate—$P(F|D)$—would have to be 19 times as large as the false positive rate—$P(F|T)$. If the likelihood ratio were less than 19, the test could not possibly lead to a situation in which the preponderance of evidence favored guilt. Thus, even if a test were perfectly accurate in detecting actual deception, a test with a false positive rate greater than .052 would not be sufficiently sensitive to detect deception that occurred in 5 cases out of 100.

Note that the conditional probabilities referred to above—$P(F|D)$ and $P(F|T)$—describe the true positive and false positive *rates,* or the proportions of deceptive and nondeceptive subjects who fail the test. These rates are always greater than or equal to the unconditional probabilities of a true positive and a true negative outcome. In particular, the unconditional probability of a false positive is given by $P(FP) = P(T) \times P(F|T)$. Thus, if the false positive rate for a test is .052, and 90% of those who take the test are telling the truth, the probability of a false positive outcome is .047.

If posterior odds of 1.0 are accepted as a lower bound for labeling a subject deceptive, the false positive rate of a test must be very low to allow any possibility of detecting infrequent deception. Assume, for example, that the test catches every person who attempts deception—$P(F|D) = 1.0$. In this case, the probability that a nondeceptive subject fails the test—$P(F|T)$—must be less than or equal to the prior odds of deception; otherwise, the test will never provide evidence that favors the verdict *guilty.*

Table 1 lists the maximum false positive rates—$P(F|T)$—that could be tolerated given this minimum threshold for guilt. The table suggests that when base rates are low (e.g., .10 or less) this maximum is approximately equal to the base rate. That is, when deception is infrequent, the tendency of the test to label innocent subjects as deceptive must be equally infrequent.

Reasonable Doubt

Assume that the posterior odds were 100 to 99 (i.e., odds of 1.01) that the subject was attempting to deceive. Although the weight of the available evidence favors a guilty verdict, the evidence is not very strong and would not be persuasive in most settings. The concept of *reasonable doubt* implies that the evidence in favor of guilt (deceptive) must, in most cases, strongly outweigh the evidence in favor of innocence (not deceptive) before a verdict of *guilty* will be pronounced. That is, the hypothesis that a subject is deceptive will not ordinarily be accepted unless the posterior odds of deception are considerably larger that one to one.

Simon and Mahan's (1971) survey of judges, juries, and lay people suggests that a legal criterion of reasonable doubt would translate into posterior probabilities of guilt between .85 and .90. Conventions for testing statistical hypotheses suggest that psychologists employ values of .95 or .99 in quantifying reasonable doubt. For the purpose of illustration, assume the criterion that a posterior probability of guilt of .90 or greater is sufficient to label a subject *deceptive.* This translates into posterior odds of 9 to 1.

Table 2 illustrates the maximum false positive rates—$P(F|T)$—that could be tolerated given this threshold for guilt. The values in Table 2 are obtained by dividing the corresponding values in Table 1 by 9, reflecting the fact that this 90% threshold for guilt requires posterior odds 9 times as large as the break-even threshold of one to one.

Table 2
Maximum False Positive Rates That Would Allow Investigator To Reach 90% Confidence Threshold for Deception

Base rate for deception	Maximum permissible false positive rate
.30	.047
.20	.027
.10	.012
.05	.005
.03	.003
.01	.001

Table 3
Maximum Permissible False Positive Rates if Test Can Be Deceived

	PDSC .98		PDSC .95		PDSC .90		PDSC .85	
Base rate	1:1	9:1	1:1	9:1	1:1	9:1	1:1	9:1
.30	.419	.046	.406	.044	.385	.042	.363	.034
.20	.245	.026	.237	.025	.225	.024	.212	.022
.10	.108	.011	.105	.014	.099	.010	.094	.010
.05	.050	.004	.049	.004	.046	.004	.044	.004
.03	.029	.002	.028	.002	.027	.002	.025	.002
.01	.009	.0009	.009	.0009	.009	.0009	.009	.0009

Note. PDSC = Proportion of deceptive subjects caught by test. Ratios represent posterior odds needed to conclude subject is deceptive.

The 90% threshold is conservative in that it implies a strong bias toward the verdict of *not guilty.* This bias may be appropriate if the test of deception leads directly to potentially adverse actions, as is frequently the case in preemployment screening. If a test of deception does not lead directly to any action, a less conservative threshold might be in order. Maximum false positive rates can be determined for any other guilt threshold by putting that threshold in odds form (e.g., an 80% confidence threshold equals odds of 4 to 1 that the person who fails the test is in fact deceptive) and dividing the values in Table 1 by those odds.

Deceiving Tests of Deception

Tables 1 and 2 are based on the assumption that tests of deception are always successful in detecting attempted deception—an assumption that is not supported by empirical research (Lykken, 1979; Sackett & Harris, 1984). When the probability of successful deception is taken into account, the limits placed by false positive rates on the detection of infrequent deception become even more extreme than those portrayed in Tables 1 and 2. Maximum permissible false positive rates, listed as a function of the base rate, the threshold for labeling a subject *deceptive,* and the probability that a deceptive individual will be caught, are shown in Table 3. The values that appear there are obtained by multiplying the corresponding values in Tables 1 and 2 by the probability of successfully detecting attempted deception—$P(F|D)$.

Several features of Table 3 are noteworthy. First, when the base rate for deception is low, the false positive rate must also be low, regardless of the threshold employed. Second, and more important, at very low base rates (e.g., .10 or lower), variations in the ability of the test to detect true deception have little impact on the overall worth of the test. For example, if the base rate for deception were .01, raising the test's probability of detecting actual deception from .85 to .98 (this would be an impressive improvement!) would have no discernible impact on the proportion of false positives that could be tolerated. If the base rate were .01, and more than 9 out of every 1,000 nondeceptive subjects failed the test, the test could not provide a preponderance of evidence to support the hypothesis that any individual examinee was deceptive. If reasonable doubt is factored in, this figure will be even lower.

Conclusion

The use of the polygraph and other similar devices for detecting and deterring deception has been the subject of intense debate in recent years. The use of these devices for screening purposes (e.g., preemployment screening) is of special concern, because in these settings the base rate for deception may be quite low. When the base rate for nontrivial deception is low (e.g., less than .10), tests must be extraordinarily sensitive to provide convincing evidence of deception. For example, if a polygraph were accurate in 98% of the cases examined (this figure is higher than the level of accuracy claimed by proponents of the polygraph, and considerably higher than the accuracy levels shown in empirical research), and errors were equally divided between false positives and false negatives, the false positive rate would be at least .0101, effectively ruling out the polygraph for base rates of .01 or lower. If a 90% reasonable doubt criterion were applied, a test with a 98% accuracy level would be too inaccurate to detect deception when the base rate is as high as .09.

It is important to note that the analyses presented here apply to any technique that might be used to detect a condition for which the base rate is low (Meehl & Rosen, 1955). For example, the Department of Defense has tested over 600,000 individuals for acquired immune deficiency syndrome (AIDS), and approximately 1.5 cases per 1,000 have tested positive ("AIDS testing screens 287," 1986). If the base rate for AIDS exposure is approximately .0015, a false positive rate as low as 2 per 1,000 would be sufficiently high to call into question the AIDS screening program.

It is not the purpose of this article to argue that tests of deception *are* useful when the base rate for deception is high; their validity is doubtful regardless of the base rate. In addition, serious ethical questions have arisen, particularly the possibility that the polygraph represents a "psychological fourth degree" that is used to extract confessions from suspects (Furedy, 1985; Furedy & Liss, 1985). Thus, even if the polygraph *were* valid, its use would be problematic. However, when the base rate is low, this analysis suggests that the debate over the value of the polygraph is meaningless because the accurate detection of deception may be beyond the capability of *any* known test or procedure.

References

AIDS testing screens 287 from military. (1986, December 13). *The Denver Post,* p. 5A.

Ben-Shakhar, G., Lieblich, I., & Bar-Hillel, M. (1982). An evaluation of polygraphers' judgments: A review from a decision theoretic perspective. *Journal of Applied Psychology, 67,* 701–713.

Brett, A. S., Phillips, M., & Beary, J. F. (1986, March 8). Predictive power of the polygraph: Can the "lie detector" really detect liars? *Lancet,* pp. 544–547.

Council takes stand on AIDS, polygraph. (1986, March). *APA Monitor,* p. 11.

Dawes, R. M., (1962). A note on base rates and psychometric efficiency. *Journal of Consulting Psychology, 26,* 422–424.

Furedy, J. J. (1985, January). Some post-Phillonic flights of polygraphic fancy. *Criminal Lawyers' Association Newsletter,* p. 3.

Furedy, J. J., & Liss, J. (1985). Countering confessions induced by the polygraph: Of confessionals and psychological rubber hoses. *Criminal Law Quarterly, 29,* 91–114.

Hollinger, R. D., & Clark, J. P. (1983). *Theft by employees.* Lexington, MA: Lexington Books.

Kleinmuntz, B. (1985). Lie detectors fail the truth test. *Harvard Business Review, 85,* 36–42.

Lieblich, I., Ben-Shakhar, G., Kugelmass, S., & Cohen, Y. (1978). Decision theory approach to the problem of polygraph interpretation. *Journal of Applied Psychology, 63,* 489–498.

Lykken, D. (1974). Psychology and the lie detector industry. *American Psychologist, 29,* 725–739.

Lykken, D. (1978). Uses and abuses of the polygraph. In H. L. Pick (Ed.), *Psychology from research to practice* (pp. 171–191). New York: Plenum Press.

Lykken, D. (1979). The detection of deception. *Psychological Bulletin, 86,* 47–53.

Lykken, D. (1981). *A tremor in the blood.* New York: McGraw-Hill.

Lykken, D. (1985). The case against the polygraph in employment screening. *Personnel Administration, 30,* 59–65.

Meehl, P. E., & Rosen, A. (1955). Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psychological Bulletin, 52,* 194–216.

Podlensy, J. A., & Raskin, D. C. (1977). Physiological measures and the detection of deception. *Psychological Bulletin, 84,* 782–799.

Raskin, D. C., & Podlensy, J. A. (1979). Truth and deception: A reply to Lykken. *Psychological Bulletin, 86,* 54–59.

Sackett, P. R., & Decker, P. J. (1979). Detection of deception in the employment context: A review and critical analysis. *Personnel Psychology, 32,* 487–506.

Sackett, P. R., & Harris, M. M. (1984). Honesty testing for personnel selection. *Personnel Psychology, 37,* 221–245.

Simon, R. J., & Mahan, L. (1971). Quantifying burdens of proof: A view from the bench, jury, and classroom. *Law and Society Review, 5,* 319–330.

U.S. Office of Technology Assessment. (1983). *Scientific validity of polygraph testing: A research review and evaluation* (Tech. Mem. OTA-TM-H-IS). Washington, DC: U.S. Government Printing Office.

Received August 5, 1986
Revision received March 23, 1987
Accepted March 26, 1987 ■