

# Thin Slices of Expressive Behavior as Predictors of Interpersonal Consequences: A Meta-Analysis

Nalini Ambady and Robert Rosenthal  
Harvard University

A meta-analysis was conducted on the accuracy of predictions of various objective outcomes in the areas of social and clinical psychology from short observations of expressive behavior (under 5 min). The overall effect size ( $r$ ) for the accuracy of predictions for 38 different results was .39. Studies using longer periods of behavioral observation did not yield greater predictive accuracy; predictions based on observations under ½ min in length did not differ significantly from predictions based on 4- and 5-min observations. The type of behavioral channel (such as the face, speech, the body, tone of voice) on which the ratings were based was not related to the accuracy of predictions. Accuracy did not vary significantly between behaviors manipulated in a laboratory and more naturally occurring behavior. Last, effect sizes did not differ significantly for predictions in the areas of clinical psychology, social psychology, and the accuracy of detecting deception.

The way in which people move, talk, and gesture—their facial expressions, posture, and speech—all contribute to the formation of impressions about them. Many of the judgments we make about others in our everyday lives are based on cues from these expressive behaviors. Gordon Allport (1937) believed that expressive behaviors were important indicators of personality and that impressions from brief interactions were often verified upon further acquaintance. Allport and Vernon (1933) demonstrated that people's expressive styles were quite consistent across a variety of situations. They then began to investigate the accuracy of perceivers' impressions that were based on observations of these expressive styles. For reasons that have been discussed elsewhere, this issue—like other issues concerning the accuracy of interpersonal and social perception—was neglected for a long time (Funder, 1987; Kenny & Albright, 1987). Recently, however, there has been a resurgence of interest in the accuracy of social and interpersonal perception and judgment (Funder, 1987; Kenny & Albright, 1987; Kruglanski, 1989; Swann, 1984) and in the study of expressive behavior (Lippa, 1983; Riggio & Friedman, 1986).

Recent work confirms earlier findings (Passini & Norman, 1966) that the ratings of strangers converge surprisingly well with self-ratings of personality by targets (Albright, Kenny, & Malloy, 1988; Funder & Colvin, 1988; Watson, 1989). This correspondence seems to confirm Gordon Allport's observation that there is something in the nature of individuals that

leads observers to attribute certain characteristics to them (Allport, 1937). We believe this "something" is communicated through expressive behavior. Much of this expressive behavior is unintended, unconscious, and yet extremely effective. For example, we communicate our interpersonal expectancies and biases through very subtle, almost imperceptible, nonverbal cues. These cues are so subtle that they are neither encoded nor decoded at an intentional, conscious level of awareness (Chaikin, Sigler, & Derlega, 1974; Christensen & Rosenthal, 1982; Harris & Rosenthal, 1985; Rosenthal, 1966; Rosenthal & Rubin, 1978; Snyder, Tanke, & Berscheid, 1977; Word, Zanna, & Cooper, 1974).

The remarkable aspect of this expressive behavior is its communicative power. A great deal of information is communicated even in fleeting glimpses of expressive behavior. Erving Goffman (1979) wrote about the "glimpsed" (p. 22) world, consisting of glimpses of strangers—a world bereft of details, yet quite rich in social information. He suggested that a rough correspondence exists between the characteristics of the people glimpsed in this world and impressions of them. Goffman used the ethological concept of "displays," or behaviors that signal inter- and intraspecies information rapidly and efficiently to explain the accuracy of impressions that are based on glimpsed behavior. He suggested that in humans, expressive behaviors constitute an important aspect of displays and, like displays in other species, these behaviors are processed naturally and efficiently by their targets.

Goffman's observations have been confirmed by subsequent research: Judgments about others can be quite accurate even when they are based on brief observations of expressive behavior (Albright et al., 1988; Funder & Colvin, 1988; Watson, 1989). One study reported that ratings that were based on 5-min videotaped clips of targets by strangers correlated to a surprising extent with the self-ratings of targets (Funder & Colvin, 1988). Others have also suggested that the unarticulated yet strong affective reactions, or "feelings," that arise from fleeting glimpses of the behavior of others might have some basis in

---

This article was part of Nalini Ambady's doctoral dissertation conducted at Harvard University. Preparation of this article was supported, in part, by a grant from the Spencer Foundation to Robert Rosenthal.

We are grateful to Roger Brown, Monica Harris, Todd Heatherton, Raj Marphatia, and five anonymous reviewers for their very helpful comments and suggestions. The content of this article is solely the responsibility of the authors.

Correspondence concerning this article should be addressed to Nalini Ambady, Department of Psychology, 1462 William James Hall, Harvard University, Cambridge, Massachusetts 02138.

reality (Schneider, Hastorf, & Ellsworth, 1979; Zajonc, 1980, 1984). This suggestion is confirmed by findings indicating that people are fairly accurate at identifying emotions from exposures to nonverbal behavior lasting only 375 ms (Rosenthal, Hall, DiMatteo, Rogers, & Archer, 1979). Other research in social and personality psychology also suggests that people might be unexpectedly accurate in the judgments they make on the basis of minimal information and minimal amounts of cognitive processing. For example, sometimes people make better judgments when they rely on their intuition rather than when they introspect or reason (Wilson & Schooler, 1991).

One way to examine the contribution of expressive behavior to the accuracy of person perception would be to rigorously control the nature of the information provided to observers and then examine the accuracy of their impressions in relation to various external criteria. This has been done by exposing observers to carefully selected clips of video- and audiotaped behavior and assessing how well their judgments predict some external criterion. A systematic review and examination of these studies would be useful in shedding light on the type of information communicated by such observations of behavior. To assess whether judgments that are based on glimpsed observations are accurate, only studies in which such observers were exposed to fairly brief segments of behavior should be considered. If fairly accurate judgments can be made solely on the basis of targets' nonverbal behavior, this result would indicate the importance of such behavior in accurate impression formation. Furthermore, the accuracy of judgments that are based on expressive behavior alone could be compared with the results of some classic studies that have examined the relationship between various types of assessment and evaluation measures and the prediction of certain outcomes. For example, in the Menninger study of psychiatric residents, a variety of psychological assessment procedures were used to predict professional competence and success, as defined by supervisor ratings and peer ratings (Holt & Luborsky, 1958). Can similar predictions be made from judgments that are based exclusively on expressive behavior?

The purpose of this article is to conduct a meta-analysis of studies on the accuracy of predictions from brief observations, or what we call "thin slices" of expressive behaviors. The results of this meta-analysis should have important theoretical and practical implications. Evidence that judgments about people on the basis of brief exposures to them are accurate would have enormous implications for the study of interpersonal perception. First, this evidence would suggest that intuitive natural judgments and perceptions of others are more accurate than one would expect. These findings would add to the body of research regarding the accuracy of day-to-day decisions (Funder, 1987; Kruglanski, 1989). Second, such evidence would also indicate that the behavior of individuals is predictable within certain situations. A third implication of these findings would be that people rapidly and unwittingly communicate a great deal of information regarding themselves to others. Fourth, this evidence would have important practical implications for various assessment, evaluation, and training procedures in clinical and other applied areas of psychology. Finally, such evidence would have important methodological implications for conducting research on expressive behavior.

## Evaluation of Accuracy

Before considering the extent of the accuracy of predictions, it is first necessary to consider how accuracy can be evaluated. The notion of accuracy implies a correspondence between a judgment and a criterion (Brown, 1965; Kruglanski, 1989). The selection of appropriate criteria to evaluate judgmental accuracy is problematic. Objective, externally valid criteria against which to evaluate predictions in the areas of social and clinical psychology are difficult to find because most of the criteria in these areas themselves often involve judgments. Addressing this issue, Kenny and Albright (1987) identified a number of criterion measures used in social perception. These measures include self-report measures, third-person (expert) judgments, objective measurements such as physiological variables, judge ratings, and operational criteria such as those used in lie-detection studies in which subjects are instructed to lie. In this meta-analysis, we considered only those studies that had (a) experimentally or objectively defined, clear behavioral criteria, corresponding to Kenny and Albright's operational criteria or (b) criteria that were ecologically valid and commonly used in everyday decisions about people, roughly corresponding to Kenny & Albright's criterion of expert judgments. We excluded self-report measures as criterion variables because they seem to be influenced by factors such as social awareness and self-knowledge of the target individual (Cheek, 1982). An example of an experimentally defined criterion would be whether a subject was actually lying in a deception-detection experiment. An ecologically valid criterion would be the use of supervisors' ratings to evaluate therapeutic effectiveness, because it is one of the primary methods of making decisions about performance and promotion. For the same reason, ratings by students would be a satisfactory criterion to evaluate college teacher effectiveness.

In the typical paradigm for research using such criteria, raters are asked to rate short samples of targets' behavior on various affective or personality dimensions. If the ratings are satisfactorily reliable, they can be used to postdict, predict, or paridict the criterion variable. For example, judges might hear short samples of therapists talking to their patients, and they might be asked to rate the therapists on a series of dimensions such as anxiety, competence, and warmth. If these ratings are reliable, they will then be correlated with the criterion variable, which might be the patient's prognosis by someone other than the therapist—such as a supervisor, or a team of caretakers, or the supervisors' ratings of the therapist. Accuracy, in this context, refers to the correspondence between the consensual judgments of the group of judges and the criterion variable.

## Accuracy From Thin Slices of Behavior: Some Findings

A number of recent studies have indicated that ratings of brief observations or thin slices of behavior can be used to predict various social and clinical psychological outcomes at levels significantly above those expected by chance (e.g. Babad, Bernieri, & Rosenthal, 1989b, 1989c; O'Sullivan, Ekman, & Friesen, 1988). We review some of the research below. This review is intended to be illustrative rather than exhaustive.

### *Clinical Outcomes*

In the clinical literature, thin slices of behavior have been used to study aspects of the therapeutic relationship. This method was advocated by Carl Rogers and his associates, who found that important variables in the Rogerian therapeutic relationship such as warmth, accurate empathy, and rapport could be satisfactorily assessed from two or three 2–5-min-segment observations of behavior. Ratings of variables from these short segments could then be used to predict patient outcomes at levels above chance (Burstein & Carkhuff, 1968; Carkhuff & Berenson, 1967; Rogers, Gendlin, Kiesler, & Truax, 1967; Truax, 1966; Truax & Carkhuff, 1967; Truax, Wittmer, & Wargo, 1971), and the reliability of ratings from short segments did not differ significantly from ratings of longer sessions (Mintz & Luborsky, 1971).

More recently, subtle expectations and biases of therapists were identified by ratings of their tone of voice while talking to and about their patients. The same ratings served to distinguish therapists with high and low ratings from their supervisors (Blanck, Rosenthal, & Vannicelli, 1986; Blanck, Rosenthal, Vannicelli, & Lee, 1986). Ratings of nonverbal behavior from brief clips have also been used to distinguish anxious and depressed people from normal people (Waxer, 1974, 1976, 1977).

### *Social Psychological Outcomes*

One area in social psychology in which thin slices of behavior have been used frequently to assess behavior with considerable accuracy is the area of interpersonal expectancies and biases. Brief clips of behavior have been used to identify successfully the subtle expressive cues conveying interpersonal expectancies that are very influential in the interpersonal influence process (Chaikin et al., 1974; Duncan & Rosenthal, 1968; Harris & Rosenthal, 1985; Rosenthal, 1966, 1969; Rosenthal & Rubin, 1978). For example, a series of studies conducted by Bugental and her colleagues revealed that parents' expectancies, identified from brief clips of their tone of voice, are related to their children's behavior (Bugental, Caporael, & Shennum, 1980; Bugental, Henker, & Whalen, 1976; Bugental & Love, 1975; Bugental, Love, Kaswan, & April, 1971). Thus ratings of the tone of voice of mothers of normal children and children with behavior problems in school differed significantly, with the latter mothers revealing a lack of confidence in their ability to control their children in their tone of voice (Bugental & Love 1975). Research in the classroom has shown that judges can distinguish biased from unbiased teachers and also can identify differential teacher expectancies and affect toward students from very brief clips of teachers' behavior (Babad, Bernieri, & Rosenthal, 1987, 1989b, 1989c). Research in the courtroom has shown that from brief excerpts of judges' instructions to jurors in actual criminal trials, raters could postdict the judges' expectations for the trial outcome and the criminal history of the defendant (Blanck, Rosenthal, & Cordell, 1985).

Ratings from thin slices have also been used to make accurate predictions regarding social outcomes pertaining to the communication of affect. For example, one set of studies investigated the communication of affect by network television newscasters. In the first study, 2.5-s long clips of the facial expres-

sions of newscasters during the 1976 presidential election campaign revealed significant differences in the facial expressions of the newscasters as a function of the candidate they were talking about (Friedman, DiMatteo, & Mertz, 1980). The second study extended this finding by relating it to an outcome measure. Ratings of 2.5-s clips of network newscasters' facial expressions during the 1984 presidential elections showed that one newscaster had significantly more positive facial expression when talking about one of the candidates. Voters who regularly watched this newscaster were significantly more likely to vote for the candidate he favored (Mullen et al., 1986).

Research on the accuracy of the detection of deception has relied almost exclusively on ratings of thin slices of behavior. A commonly used experimental paradigm involves subjects being audiotaped or videotaped while honestly describing someone they like, honestly describing someone they dislike, and dishonestly pretending to like the disliked target and dislike the liked target. From short clips of their behavior in each of these conditions, judges are asked to rate the honesty of the subject (e.g., DePaulo & Rosenthal, 1979). Reviews of research on the accuracy of deception detection have found that the mean accuracy of detection was above chance (DePaulo, Zuckerman, & Rosenthal, 1980; Zuckerman, DePaulo, & Rosenthal, 1981), although it appears that it is influenced by a number of factors in the experimental situation (Bond, Kahler, & Paolicelli, 1985; DePaulo, Kirkendol, Tang, & O'Brien, 1988).

How well do ratings from thin slices of behavior predict clinical and social outcomes compared with predictions that are based on other methods? This question is addressed later in this article. Although short segments of behavior apparently can be used to predict socially and clinically important outcomes successfully, there has been some controversy about whether ratings of specific behavioral channels (such as speech, posture, or facial expression) are linked to more accurate assessment (Archer & Akert, 1977; Argyle, Alkema, & Gilmour, 1971; Brown, 1986, pp. 496–502; Mehrabian & Wiener, 1967). We review some of the pertinent research in the next section.

### Accuracy of Predictions From Different Channels of Communication

Studies on the accuracy of predictions from short segments of behavior have varied the behavioral channels shown to judges. The typical channels shown include (a) the nonverbal channels, which include the visual channels (face, body, or face and body) and the vocal channel (just tone of voice), (b) the verbal channels, which include speech and transcripts, and (c) the audiovisual channel (combining the visual and verbal channels). In comparing these channels, most of the research has addressed two related issues. The first concerns the relative efficacy of the verbal versus the nonverbal channels. The second concerns the relative efficacy of the various nonverbal channels.

Most of the work reviewed in this section has been drawn from the social psychological literature, especially work on the detection of deception. Studies in the clinical literature have indicated that ratings do not seem to differ significantly across channels (Burns & Beier, 1973; English & Jelenevsky, 1971;

Strahan & Zytowski, 1976; Strong, Taylor, Bratton, & Loper, 1971).

Goffman (1959, 1971) pointed out that expressive behaviors can be controlled to meet certain self-presentation goals and to convey certain impressions. Although we are generally able to monitor and control aspects of our behavior according to the "display rules" that determine what behaviors are culturally and socially appropriate, sometimes our true feelings "leak" out through the behavioral channels that are less controllable (Ekman & Friesen, 1969). At some level, perceivers realize that people can choose their words carefully, but they are less adept at controlling their facial, vocal, and bodily expressions. The lack of control of nonverbal behavior could be attributed to a lack of awareness of these behaviors, because people cannot see or hear themselves as others do.

The most controllable channel is the verbal channel (speech), followed by the face, the body, and the least controllable channel, the voice (Brown, 1986; Rosenthal & DePaulo, 1979). The leakier, less controllable, nonverbal channels should be more accurate. When verbal and nonverbal behaviors are inconsistent, nonverbal behaviors may be more revealing of the true message. For example, Blanck et al. (1985) found that judges who expected a defendant to be guilty revealed their expectation in their nonverbal behavior, but not in their verbal behavior. Similarly, distressed married couples trying to act happy could be distinguished from happy couples by their nonverbal behavior rather than their verbal behavior (Vincent, Friedman, Nugent, & Messerly, 1979). Much of the research comparing channels of communication, however, has focused on the accuracy of identifying deceptive or inconsistent messages. In a theoretical review of the literature, Noller (1985) found that access to the leaky channels becomes more important for accuracy for deceptive or inconsistent behaviors. But meta-analytic results clearly indicate that the presence of verbal content improves the accuracy of detecting deception (DePaulo et al., 1980; Zuckerman, DePaulo, & Rosenthal, 1981). Although the issue of the relative importance of the verbal and the nonverbal channels has not been resolved, the relative importance of the channel apparently depends on a number of different factors. These include (a) the expectations of the judges (Zuckerman, Spiegel, DePaulo, & Rosenthal, 1982), (b) the type of message being conveyed (Apple, Streeter, & Krauss, 1979; Ekman, 1988; Ekman, Friesen, O'Sullivan, & Scherer, 1980; Streeter, Krauss, Geller, Olson, & Apple, 1977; Zuckerman, Amidon, Bishop, & Pomerantz, 1982; Zuckerman, Larrance, Spiegel, & Klorman, 1981), (c) situational factors such as familiarity with the situation regarding which judgments have to be made (Krauss, Apple, Morency, Wenzel, & Winton, 1981; Stiff et al., 1989; Zuckerman, Spiegel, DePaulo, & Rosenthal, 1982), (d) the type of affect being expressed and the type of affect being rated (DePaulo & Rosenthal, 1979; DePaulo, Rosenthal, Eisenstat, Rogers, & Finkelstein, 1978; Noller, 1985; Scherer, Scherer, Hall, & Rosenthal, 1977; Zuckerman, Hall, DeFrank, & Rosenthal, 1976; Zuckerman, Spiegel, DePaulo, & Rosenthal, 1982), (e) the quality of the information being transmitted by the various channels (Argyle et al., 1971; Gallios & Callan, 1986), and (f) the motivation of the subjects (DePaulo et al., 1988; DePaulo, Lanier, & Davis, 1983).

What implications do these findings regarding the contribu-

tions of the various behavioral channels have for this article? We expect that the accuracy of predictions from brief observations of behavior will not be significantly different for the different channels. If anything, information from observations including verbal information should be superior. In summary, the major questions addressed by this meta-analysis are (a) can accurate predictions be made from short observations of expressive behavior, (b) if they can be made, how accurate are these predictions, and (c) are certain behavioral channels associated with more accurate predictions from thin slices of behavior?

## Method

### Literature Search

Four methods were used to locate the relevant studies. First, an initial computer search of *Psychological Abstracts* was conducted to retrieve documents containing the terms *accuracy*, *deception*, and *nonverbal behavior*. However, because our main criterion for inclusion of studies in this analysis was a methodological one, this method did not prove to be very useful. The second method was a manual search of volumes of the following journals covering a time span ranging from 1970 to 1990. Journals were selected on the basis of relevance to social and clinical psychology and citation in other relevant articles and books. Journals manually searched for articles included the *European Journal of Social Psychology*, *Journal of Abnormal Psychology*, *Journal of Applied Social Psychology*, *Journal of Clinical Psychology*, *Journal of Communication*, *Journal of Consulting and Clinical Psychology*, *Journal of Counseling Psychology*, *Journal of Educational Psychology*, *Journal of Experimental Social Psychology*, *Journal of Nonverbal Behavior*, *Journal of Personality and Social Psychology*, and *Personality and Social Psychology Bulletin*. Third, publications (especially those published before 1970) were located by searching reference lists of relevant articles and books. Last, our own files provided preprints and unpublished manuscripts pertinent to this investigation.

The last three methods yielded nearly 100 studies. To be included in this review, a study had to meet the following criteria:

1. Ratings or judgments should have been based on no more than 300 s of behavioral observations. Every observation of the same subject was included in estimating the observation length. For example, if sixty 10-s clips of the same subject were rated in a study, judges would have observed the subject for 600 s, and the study would have been excluded from this analysis. If no information was provided on the length of observations and if length could not be assessed from other information provided, the study was not included.<sup>1</sup> We chose a cutoff of 5 min rather arbitrarily, in appreciation of the work by Pittenger, Hockett, and Danehy (1960), who elegantly described the richness of information that can be communicated in just 5 min in the context of a diagnostic therapeutic interview. Although subsequent work has shown that a great deal of information can be conveyed in even briefer time periods (e.g., DePaulo & Rosenthal, 1979; Milmoie, Novey, Kagan, & Rosenthal, 1968; Milmoie, Rosenthal, Blane, Chafetz, & Wolf, 1967; Rosenthal, Blanck, & Vannicelli, 1984), we selected 5 min (300 s) as our cutoff for this review.

2. Short behavioral ratings had to be related to some clearly defined external, objective, behavioral criterion or to the criterion of ratings by

<sup>1</sup> Some studies did not provide the observation length, but it was possible to estimate it from other information provided in the study. For instance, one study gave examples of four scenes typical for length and message type (Bugental, Love, Kaswan, & April, 1971). We had 3 people read these messages at a normal rate of speaking and averaged their timed responses to get an estimate of observation length.

experts. An external criterion would be the existence of deception. An expert criterion would be supervisor's ratings of effectiveness. Self-report ratings were not used as criterion variables nor were peer ratings.

3. The study had to contain enough information to permit estimation of the significance level and effect size.<sup>2</sup> Forty-four studies met all the criteria for inclusion. These studies certainly represent a large sample of the population of studies that met all our criteria, although it is always possible that some studies might have been overlooked. The potential impact of studies we might have missed will be considered in the Results section. While coding all the studies, it became clear that a few studies were correlated (because they had used the same subjects). Results from these were combined, and the final sample consisted of 38 independent results drawn from 44 studies. Table 1 contains information about the authors, year of publication, criterion variable, and total length of observation for each study.

### Coding Procedure

The following variables regarding subjects and raters or judges were coded for each study: (a) number of subjects overall (stimulus persons or ratees), (b) proportion of female subjects, (c) whether the subjects were college students, non-college students, or children, (d) whether behavior of the subjects was naturally occurring or experimentally manipulated, (e) the number of raters (or judges), and (f) the proportion of female raters.

General information about the study recorded included (a) whether it was a field or a laboratory study (this category overlapped considerably but not completely with the previously mentioned category regarding naturally occurring or manipulated behavior), (b) whether the outcome variables could be defined as related to clinical psychology, social psychology or the psychology of deception (although studies on deception fall under the general rubric of social psychology, they were considered a separate category because of previous work on deception accuracy as a separate category), (c) the year of publication of the study, and (d) the publication outlet.

Specific information coded for each study included (a) the number of clips rated for each subject, (b) the number of behavioral clips rated by each judge, (c) the length of each clip, (d) the length of the total observation time for each subject, (e) whether nonverbal behavior alone or whether both verbal and nonverbal behaviors were rated, and (f) the behavioral channels rated (face, body, speech, tone of voice, transcripts, or combinations of the above).

Information about the results also included the magnitude (the effect size) and the significance level of the relationship between the criterion variable and the behavioral dimensions rated. For each study, only one effect size and one level of significance were recorded to meet the requirement of independence of effect sizes and significance levels. When there were multiple results that seemed to be correlated in a single study, and the correlation between these multiple dependent variables could be estimated, Rosenthal and Rubin's (1986) formula was used to compute effect sizes and significance levels.<sup>3</sup> When the intercorrelation could not be estimated, the mean of the relevant results was used in the meta-analysis. This is a robust procedure, although it is conservative, and probably deflated the effect sizes (Rosenthal, 1984; Rosenthal & Rubin, 1986).<sup>4</sup> Also coded were (a) the direction of the effect and (b) the effect sizes and significance levels separately for each behavioral channel, if they were reported separately in the studies. The meta-analytic procedures used to analyze these data are those described in Rosenthal (1984).<sup>5</sup> The effect size ( $Zr$ ) and level of significance ( $Z$ ) for each study are also presented in Table 1.

### Study Characteristics

Of the 44 studies, 68.2% (30) were reported between 1980 and 1990, 27.3% (12) between 1970 and 1979, and 4.5% (2) between 1960 and

1969. The median year of publication was 1983. Forty-one studies appeared in journals, 1 was reported in a book chapter, and 2 were unpublished manuscripts. The median number of subjects was 32 and the range was from 2 to 271. Of the 38 results analyzed, 6 studies used only female subjects, 8 used only male subjects, 9 used equal numbers of male and female subjects, 2 used between 51% and 99% female subjects, 2 between 51% and 99% male subjects, and 11 did not report subject gender. The median number of judges per study was 37, and the number ranged from 2 to 446. All the judges were naive subjects. Five studies reported using only female judges, 1 study used only male judges, 13 studies used equal numbers of male and female judges, 3 studies used 51–99% female judges, 3 studies used 51–99% male judges, and 13 did not report genders.

## Results

Table 2 contains a stem and leaf display of the effect sizes of the studies in the meta-analysis, Table 3 contains some additional useful information about central tendency, variability, significance tests, and confidence intervals.

### Effect Size and Significance Testing

All the results show a positive effect size for accuracy in predictions, where 50% would be expected under the null hypothesis. The mean effect size for accuracy of judgment from all segments of behavior under 300 s was .39. This effect was asso-

<sup>2</sup> In a critique of conclusions on the accuracy of deception detection (e.g., DePaulo, Zuckerman, & Rosenthal, 1980), Kraut (1980) argued that hit rates should be considered in addition to effect sizes in estimating accuracy. If looked at in this way, accuracy scores for the detection of deception rarely exceed 65%, which is regarded as a medium effect size by Cohen (1988), given a chance level of 50%. Although we agree that providing several effect-size estimates can be useful in some circumstances, our analysis was confined to the more common effect sizes, largely because most of the studies reviewed provided no information regarding hit rates.

<sup>3</sup> For example, in one study, therapists' tone of voice in talking about patients was used to predict how they would talk to those patients. Judges rated the latter dependent variable on dimensions of warmth, dominance, empathy, competence, hostility, anxiety, optimism, professionalism, honesty, and liking (Rosenthal, Blanck, & Vannicelli, 1984). The typical correlation among these variables is .50; this was the value entered into Rosenthal and Rubin's (1986) equation. The purpose of this equation is to get a more accurate estimate of effect sizes and significance levels, because traditional methods such as computing the mean or median are too conservative (see Rosenthal & Rubin, 1986, for more information).

<sup>4</sup> To estimate the mean effect size, each  $r$  value was transformed to  $zr$  before computing the mean, which was then converted into the  $r$  for this study. Estimation of the mean significance level was done in one of two ways: (a) When the  $r$  was reported, the  $t$  for each result was computed from the  $r$  using the formula  $t = r[d/f/(1 - r^2)]^{1/2}$ ; the corresponding  $Z$  (standard normal deviate) associated with the  $p$  value of each  $t$  was found and these  $Z$ s were averaged, and the  $p$  level corresponding to the average  $Z$  was the significance level of the study, and (b) when the  $r$  was not reported, the  $Z$  was found for the  $p$  value reported in the result, and  $r$  was obtained from the formula:  $r = Z/N^{1/2}$ . If results were reported as nonsignificant and no  $p$  value was specified, the  $p$  value was assumed to be .50 with a  $Z$  of 0 (Rosenthal, 1984).

<sup>5</sup> We used a program written by Monica Harris (1984) for the meta-analytic computations.

Table 1  
*Summary Table of Studies, Criterion Variables, and Length of Exposure*

Authors	Area	Criterion variable	Length (in seconds)	Zr	Z
Ambady and Rosenthal (1990)	S	Teacher effectiveness	30	0.95	3.62
Apple and Hecht (1982)	S	Actual affect of subjects	96	0.23	3.48
Archer and Akert (1977)	S	Actual social behavior of subjects	30-60	0.83	12.18
Babad, Bernieri, and Rosenthal (1989b)	S	Existence of bias in teachers	270	0.60	3.16
Babad, Bernieri, and Rosenthal (1989a)	S	Expectancies of teachers	120	0.42	5.24
Babad, Bernieri, and Rosenthal (1987)	S	Status of teachers	270	0.55	3.12
Blanck, Rosenthal, and Vannicelli (1986) <sub>a</sub>	C	Status of patients	50-100	0.15	3.32
Blanck, Rosenthal, Vannicelli, and Lee (1986) <sub>a</sub>	C	Supervisor ratings of therapists	50-100		
Rosenthal, Blanck, and Vannicelli (1984) <sub>a</sub>	C	Way therapist would speak to patients	30-60		
Bugental, Love, Kaswan, and April (1971)	C	Types of families	3.5	0.16	0.85
DePaulo and Rosenthal (1979) <sub>b</sub>	D	Existence of deception and type of affect	240	0.28	4.26
DePaulo, Rosenthal, Green, and Rosenkrantz (1982) <sub>b</sub>	D	Existence of deceptive and mixed messages	120		
DePaulo, Lanier, and Davis (1983)	D	Existence of deception	120	0.59	4.37
DePaulo, Lassiter, and Stone (1982)	D	Existence of deception	120	0.16	1.82
DePaulo, Stone, and Lassiter (1985)	D	Existence of deception	300	0.27	1.95
Feldman (1976)	D	Existence of deception and honesty in teachers	30	0.33	2.00
Feldman, Jenkins, and Popoola (1979)	D	Existence of deception	20-40	0.44	2.29
Feldman and Prohaska (1979)	S	Adequacy of teachers	20	0.37	2.21
Friedman, Hall, and Harris (1985)	C	Score on index of peripheral artery disease	35	0.27	3.20
Fugita, Hogrebe, and Wexley (1980)	D	Existence of deception	240	0.21	3.75
Hall and Braunwald (1981)	S	Detecting gender of target being addressed	40	0.42	3.08
Hall, Roter, and Rand (1981)	C	Predicting patient commitment & compliance	90	0.30	2.94
Hall, Roter, and Katz (1987) <sub>e</sub>	C	Physician proficiency and patient satisfaction	90	0.22	3.72
Roter, Hall, and Katz (1987) <sub>e</sub>					
Kaul and Schmidt (1971)	C	Trustworthiness in interviewers	300	0.47	2.32
Machida (1986)	S	Comprehension in children	120	0.93	4.75
Manstead, Wagner, and MacDonald (1986)	D	Existence of deception	60	0.32	1.82
Milmo, Rosenthal, Blane, Chafetz, & Wolf (1967)	C	Referral of alcoholic patients	90	0.42	1.64

Table 1 (continued)

Authors	Area	Criterion variable	Length (in seconds)	Zr	Z
Milmoie, Novey, Kagan, & Rosenthal (1968)	C	Behavior of babies	51	0.42	1.70
Mullen et al. (1986)	S	Voting behavior	25	0.23	2.55
O'Sullivan, Ekman, and Friesen (1988)	D	Existence of deception	120	0.14	0.78
Riggio and Friedman (1983)	D	Existence of deception	78	0.29	3.80
Riggio, Tucker, and Throckmorton (1987)	D	Existence of deception	120	0.11	1.14
Steckler and Rosenthal (1985)	S	Communication with boss, peer, or subordinate	30	0.11	0.91
Stiff et al. (1989)	D	Existence of deception	180	0.25	5.12
Streeter, Krauss, Geller, Olson, and Apple (1977)	D	Existence of deception	26	0.21	2.42
Waxer (1977)	C	High- versus low- anxiety patients	60	0.58	3.25
Waxer (1976)	C	Depression in patients	120	0.58	4.08
Waxer (1974)	C	Depression in patients	120	1.33	9.59
Zuckerman, DeFrank, Hall, Larrance, and Rosenthal (1979)	D	Existence of deception	120	0.60	8.17
Zuckerman, Koestner, Colella, and Alton (1984) <sub>d</sub>	D	Existence of deception	200	0.74	13.26
Zuckerman, Koestner, and Alton (1984) <sub>d</sub> <sup>a</sup>	D	Existence of deception	200		
Zuckerman, Fisher, Osmun, Winkler, and Wolfson (1987) <sub>d</sub>	D	Existence of deception	200		
Zuckerman, Driver, and Guadagno (1985)	D	Existence of deception	100	0.26	2.35

*Note.* Length of the clips was calculated by combining the length of clips for each subject. Thus, if judges rated three 10-s clips of the same subject, length would be recorded as 30 s. Studies having the same subscript were combined in this analysis. C = studies with clinical criteria, D = studies on accuracy of deception, S = studies with social psychological criteria.

<sup>a</sup> Results from the control condition alone were included in this analysis, because the experimental conditions were training conditions to increase the accuracy of detecting deception.

ciated with a statistically significant  $Z$  of 22.56,  $p < .1^{112}$ . When weighted by the degrees of freedom, the mean  $r$  was .41. This significant and substantial effect size (Cohen, 1988) indicates that people can predict outcomes quite accurately from very small segments of behavior. The 95% confidence interval suggests the likely range of effect sizes to be from .34 to .48.

The coefficient of robustness (Rosenthal, 1990) is the reciprocal of the coefficient of variation and provides an index of the stability and replicability of the average effect size. It does not increase with the increasing number of replications. Robustness increases as the variance in effect sizes decreases and as the distance of the mean effect size from zero increases. Although comparative data do not exist yet, we hope that researchers will report this statistic in the future.

Of the 38 results analyzed here, 11 were coauthored by Robert Rosenthal, 10 were authored by Rosenthal's former students, and 17 were from other laboratories. We compared these results and found for studies coauthored by one of us the effect size  $r$

was .38 ( $Z = 10.79$ ), for studies authored by Rosenthal's former students the effect size  $r$  was .40 ( $Z = 16.08$ ), and for studies from other laboratories the effect size  $r$  was .39 ( $Z = 12.71$ ). These effect sizes were very homogeneous,  $F(2, 35) = .04$ ,  $p > .1$ .

#### *File Drawer Analysis*

Results that fail to reach statistical significance are less likely to have been published (Rosenthal, 1984). A number of such studies might have accumulated in researchers' file drawers. For the probability for this meta-analysis to become nonsignificant ( $p > .05$ ), using simple calculations (Rosenthal, 1984, 1987) there would have to be 7,110 studies with mean probability of .50 languishing in file drawers.

#### *Additional Analyses*

Correlational analyses revealed that neither the number of judges or subjects nor their gender was significantly related to

the accuracy of predictions. The latter result was surprising because previous studies indicated that women tended to be more accurate interpreters of nonverbal cues than men (Hall, 1984). But this superiority is also found to depend on the type of cue; women are better at decoding the less but not the more "leaky" channels (Rosenthal & DePaulo, 1979). Because this meta-analysis included studies using all the various channels, the superiority effect for women might have been canceled out.

For further analyses, studies were classified and compared along various categories of interest described below. These categories and their respective results are displayed in Table 4.

*Length of exposure.* A central issue in this review was whether increasing the length of the audio or video clips rated would increase the accuracy in predicting criterion variables. One of our most important findings in terms of its methodological implications was that effect sizes for studies with varying length of exposures (less than 30 s, 30–60 s, 60–120 s, 120–180 s, 180–240 s, or 240–300 s) were very close, ranging from .24 to .45 (all associated with statistically significant *Z*s). A linear contrast testing the effect of exposure length on mean effect sizes was not significant (*Z* = .03), indicating that accuracy does not increase with longer exposures. A second contrast compared the effect size for exposures under 30 s to the other five effect sizes. This contrast was also not significant (*Z* = -.11). This suggests that judgments from very brief segments of behavior (under half a minute in length) may be as accurate as judgments from longer segments (up to 5 min long). Longer exposures do not seem to increase accuracy significantly.

*Type of outcome.* Classification of studies by the type of outcome predicted revealed an average effect size of .41 for studies with clinical outcomes. Studies with general outcomes in the area of social psychology had an average effect size of .47<sup>6</sup> and the effect size for those with outcomes specifically related to the accuracy of detecting deception was .31. On the whole, the type of outcome predicted apparently is not significantly related to the accuracy of predictions, as shown by the results of a linear contrast done on the effect sizes (*Z* = -.26; weights for the contrast of -1 for clinical outcomes, 0 for social, and 1 for deception studies reflected the amount of experimental control over the stimuli).

*Type of study.* Combined effect sizes obtained for studies in which subjects' behavior had been experimentally manipulated in the laboratory were compared with those obtained from

Table 3  
Statistical Summary of 38 Results (45 Studies)

Statistic	Value
Central tendency ( <i>r</i> )	
Unweighted <i>M</i>	.39
Weighted <i>M</i>	.41
Proportion >0.00	100.0%
Significance tests	
Combined Stouffer <i>Z</i>	22.56
<i>t</i> test <sup>a</sup> for mean <i>r</i>	11.39
Variability ( <i>r</i> )	
Maximum	.87
Quartile 3 (Q3)	.52
Median (Q2)	.30
Quartile 1 (Q1)	.22
Minimum	.10
Q3 - Q1	.30
$\hat{\sigma}$ [.75(Q3 - Q1)]	.23
<i>s</i>	.19
SD/ $\sqrt{N}$ (SE)	.03
Robustness ( <i>M</i> / <i>SD</i> )	2.01
CI for <i>r</i> <sup>b</sup>	
95%	.34-.48
99%	.31-.51
99.9%	.28-.54

<sup>a</sup>  $t = \frac{\bar{Z}r}{\sqrt{1/38(SD^2)}}$     <sup>b</sup> CI = confidence interval (*n* = 38).

nonexperimental studies in which subjects' behavior was allowed to vary naturally. A contrast on the effect sizes (experimental studies *r* = .32, field studies *r* = .47) showed that the degree of control used in the study was not significantly related to the accuracy of predictions. Note that most of the experimental studies were in the area of detecting deception, and most of the nonexperimental studies related to the prediction of clinical criterion variables. As far as they go, then, these results do not overwhelmingly support the proposal that observations from natural situations should be more accurate than observations from laboratory situations (Funder, 1987), although it is interesting that the effect size for field studies was higher than the effect size of laboratory studies.

*Type of behavior.* For this analysis, results were categorized according to whether only nonverbal behavior was rated or whether verbal behavior was also included (usually speech or transcripts). Again a contrast on the effect sizes indicated that there were no significant differences between the two categories (the correlation for nonverbal alone was .45; for verbal and nonverbal, the correlation was .35). Both effect sizes were associated with significant *Z*s.

Results for Different Channels

Table 5 contains the results for different channels rated across all studies. There are 65 results because in some studies

<sup>6</sup> This category included 6 studies related to the social psychology of education. The *Z* for the education-related studies was 9.20, with a corresponding effect size *r* of .56; *Z* for the 5 studies with pure social psychological outcomes was 9.93, and the *r* was .35. These results were not significantly different (*Z* = -.29), so they were combined under the category of social psychology.

Table 2  
Stem and Leaf Plot of Mean Effect Size (*r*)

Stem	Leaf
1.0	
.9	
.8	7
.7	3, 4
.6	3, 8
.5	0, 2, 2, 3, 4, 4
.4	0, 0, 0, 1, 7
.3	1, 3, 5
.2	1, 1, 1, 2, 3, 3, 4, 5, 6, 6, 7, 8, 9
.1	0, 0, 4, 5, 6, 6
.0	



Table 4  
Results of Studies Classified by Length of Exposure, Type of Outcome, Type of Study, and Type of Behaviors Rated in the Study

Category	n	Z	r	Trimmed		Weighted r	SD <sup>a</sup>	Rob
				Z	r			
Exposure length <sup>b</sup>								
0-30	7	5.50	.33	4.51	.26	.25	.21	1.57
30-60	7	10.38	.44	6.10	.41	.57	.14	3.14
60-120	16	15.03	.39	13.35	.35	.36	.23	1.70
120-180	1	5.12	.24	—	—	—	—	—
180-240	3	12.28	.39	4.26	.27	.40	.23	1.70
240-300	4	5.28	.45	3.84	.49	.39	.13	3.46
Type of outcome								
Clinical	11	10.69	.41	8.44	.35	.42	.22	1.86
Deception	16	14.83	.31	12.09	.30	.31	.15	2.07
Social	11	13.36	.47	10.40	.47	.52	.21	2.24
Type of study								
Field	17	13.89	.47	12.17	.44	.44	.21	2.24
Laboratory	21	17.85	.32	15.55	.32	.40	.16	2.00
Behaviors rated								
NV	15	14.42	.45	12.68	.42	.39	.24	1.88
V + NV	23	17.35	.35	15.09	.34	.42	.16	2.19

Note. Rob = coefficient of robustness, NV = nonverbal, V = verbal. Dashes signify nonapplicable.

<sup>a</sup> Based on *r*. <sup>b</sup> In seconds.

ratings were made separately for different channels. However, only one result was coded per channel for each study. The effect sizes ranged from .26 for ratings that were based on transcripts to .54 for ratings that were based on the face and body. All effect sizes were associated with significant *Z*s, and linear contrast analyses revealed no significant differences between them—contrasts compared the individual channels (face, body, and speech) to the combined channels, because the effect for the combined channels was expected to be larger. Judges apparently were most accurate when they were able to observe the face and the body ( $r = .54$ ). Across all studies, the level of accuracy declined to .28, though not significantly so ( $Z = -.72$ ), when speech was added to the face and body. However, this issue could also be looked at within studies. In three studies, some judges saw just the face and body whereas others were exposed to the face, body, and speech; these studies found that ratings that were based on the face, body, and speech ( $r = .42$ ) were more accurate than ratings that were based on just the

Table 5  
Accuracy of Judgments for Different Channels

Channel	n	Z	r
Body	2	2.32	.28
Face	5	7.32	.40
Speech	8	8.20	.36
Tone of voice	12	7.66	.26
Transcripts	6	4.75	.29
Body + speech	2	2.92	.33
Face + body	12	16.24	.54
Face + speech	3	9.17	.41
Face, body, and speech	15	9.54	.28

face and the body ( $r = .24$ ), although the two were not quite significantly different ( $Z = 1.56$ ). Excluding the studies in which ratings had been obtained on the basis of the face and body both with and without speech, ratings of just the face and body were more accurate ( $r = .62$ ) than those based on the face, body, and speech ( $r = .24$ ), although again the difference was not significant ( $Z = -.91$ ), suggesting that too much information was confusing or distracting to the judges. For the individual channels, ratings that were based on the face alone ( $r = .40$ ) or on speech alone ( $r = .36$ ) were slightly though not significantly more accurate than ratings that were based just on transcripts ( $r = .29$ ), the body ( $r = .28$ ), or tone of voice ( $r = .26$ ).

Previous reviews focusing only on the accuracy of detection of deception had indicated that verbal content improved accuracy considerably (DePaulo et al., 1980; Zuckerman et al., 1981). These reviews included some studies with brief clips. To find out whether the presence of substantive verbal content increased accuracy in the present review, results were combined across channels for the presence or absence of speech. These are shown in Table 6.

The presence or absence of speech apparently does not significantly affect judgmental accuracy when exposures to stimuli are brief. The difference between our results and the results of the previous meta-analyses may be due to the fact that the majority of the studies included in the latter used longer clips of behavior. When observations are longer, cues from the verbal channel might be more informative and lead to greater prediction accuracy.

### Comparisons Within Areas

We also examined the effect of length of observation, the behaviors rated, and the type of study within the clinical, deception, and social studies. These results are presented in Table 7.

Note that there are missing data for some of the columns, and some of the results are based on single studies. In the clinical area, there were no significant differences between studies with differing lengths of exposure (contrast  $Z = .08$ ) or different channels (contrast  $Z = -.17$ ). Within the social area, there were no differences between various lengths of exposure ( $Z = .04$ ), channels ( $Z = -.07$ ), or the type of study ( $Z = -.19$ ). Similarly, for deception studies, there were no significant differences in exposure lengths ( $Z = .06$ ) or different channels ( $Z = .05$ ).

Table 6  
Presence of Content and Accuracy of Judgments for Different Channels

Content	Channel			<i>M</i>
	Face (F)	Body (B)	F + B	
Present (speech)	.41	.33	.28	.34
<i>N</i>	3	2	15	
Absent (no speech)	.40	.28	.54	.41
<i>N</i>	5	2	12	
<i>M</i>	.41	.31	.41	.38

Table 7  
*Comparing Length of Exposure, Type of Study, and Type of Behavior  
 Rated Within Different Outcome Areas*

Category	Outcome variable								
	Clinical			Social			Deception		
	<i>n</i>	<i>Z</i>	<i>r</i>	<i>n</i>	<i>Z</i>	<i>r</i>	<i>n</i>	<i>Z</i>	<i>r</i>
Exposure length*									
0-30	1	0.85	.16	4	4.65	.39	2	3.13	.27
30-60	3	4.67	.40	2	10.79	.56	2	2.91	.36
60-120	6	9.88	.43	3	7.78	.48	7	8.48	.30
120-180	—	—	—	—	—	—	1	5.12	.24
180-240	—	—	—	—	—	—	3	12.28	.39
240-300	1	2.32	.47	2	4.40	.52	1	1.95	.26
Type of study									
Field	10	10.03	.43	7	9.65	.53	—	—	—
Laboratory	1	3.72	.22	4	9.39	.36	16	14.38	.31
Behavior rated									
NV	7	9.88	.46	5	7.43	.50	3	7.57	.32
V + NV	4	4.67	.30	6	11.30	.45	13	12.81	.31

Note. Dashes signify nonapplicable. NV = nonverbal, V = verbal.

\* In seconds.

### *Thin Compared With Thick Slices of Behavior*

At this stage, the reader might wonder how predictions from thin slices of behavior compare with those made from other sources. We could not locate studies that have compared predictions from brief clips directly with predictions from other sources. However, a few studies have used other methods to predict criterion variables similar to the ones predicted from brief clips. Wiggins (1973) discusses some of the classic American "milestone" studies on predicting criterion variables by various evaluation and assessment methods—including interviews, self-report, and projective personality assessment measures. We compared the results of this meta-analysis with the results of the classic milestone studies reported by Wiggins as well as with a few other studies that have used criterion measures similar to ours.

Our first comparison was with Wiggins's (1973) summary of the results of the Office of Strategic Services (OSS) assessment study. The OSS study was undertaken in 1948 in an effort to predict the effective performance of OSS officers. Predictions were based on a variety of psychological tests, situational performance measures, and interviews. The criterion variables included appraisals on the job by superiors and assessment staff. Wiggins reported the correlations between various assessment and appraisal ratings (Table 11.4, p. 534). We combined the correlations and calculated that the average correlation between predictor and criterion variables was .26. Contrasting this effect size to the effect size of the present study ( $r = .39$ ) revealed no significant difference in the accuracy of prediction between the two studies ( $Z = .62$ ). Results comparing the effect size obtained in the present meta-analysis with results from other studies are displayed in Table 8.

Our second comparison was with the results of another milestone study reported by Wiggins (1973). This is the Veterans Administration (VA) assessment project conducted between

1946 and 1949 to evaluate procedures used to select clinical psychologists. Clinical trainees were evaluated at various testing centers by staff judges using a variety of psychological tests and interviews, and these assessments were related to a number of different criterion measures including performance ratings from university departments, analyses of examination performance, field tests of work samples, and ratings by supervisors and colleagues. Again, our computations are based on results reported by Wiggins (Table 11.11; p. 563). Results comparing the average correlation obtained in the VA study from the pooled assessment of the judges and from the best psychological test predictors are reported separately in Table 7. Again, contrast analyses revealed that these effect sizes did not differ significantly from the effect size obtained in the present study.

Our third example is also with research in clinical psychology and is based on the work of Holt and Luborsky (1958). This study is one of the milestone studies reported by Wiggins (1973), but we compared our results with those from a table not included in his write-up of the study. Holt and Luborsky studied over 200 psychiatric residents at the Menninger School of Psychiatry, using several different methods to predict psychiatric competence. After considerable deliberation, they selected supervisors' evaluation of overall competence as the major criterion variable. Another criterion variable they used was peer ratings of competence. Our comparison is based on the third of their three studies, which cross-validated predictors from the previous two studies. Four judges' ratings of about 65 residents, using different methods of evaluation (such as analyses of application materials, interviews, Thematic Apperception Test, and Rorschach protocols), were correlated with peer and supervisor ratings of residents' competence. Judge ratings were correlated with supervisor and peer ratings of the residents on psychotherapy competence, diagnostic competence, management competence, and overall competence. The average effect size relating judges' overall ratings and the criterion variable of supervisor

Table 8  
*Contrasting Effect Sizes (r) That Are Based on "Thick" Slices of Behavior  
 With Effect Sizes From the Present Study*

Predictor variable	Criterion variable	Study <i>r</i>	Contrast <i>z</i>	Contrast <i>p</i>
OSS study <sup>a</sup> appraisal ratings	Job performance	.26	0.62	.27
VA assessment study <sup>b</sup> Best objective test	Clinical competence	.28	0.56	.29
		.29	0.56	.29
Menninger study <sup>c</sup> Judges' ratings and psychological tests	Clinical competence	.31	0.17	.44
Meta-analysis <sup>d</sup> Self-reports	Teaching effectiveness (students' ratings)	.07	1.36	.09
	(students' ratings based on classroom performance)	.41	0.06	.47
		.33	0.19	.42
Meta-analysis <sup>e</sup> Judges' ratings	Deception detection	.32	0.12	.45
Meta-analysis <sup>f</sup> Judges' ratings	Deception detection	.45	0.54	.29
Median <sup>g</sup>		.31		

*Note.* OSS = Office of Strategic Services, VA = Veterans Administration.

<sup>a</sup> Wiggins (1973), p. 534. <sup>b</sup> Wiggins (1973), p. 563. <sup>c</sup> Holt and Luborsky (1958), p. 213. <sup>d</sup> Feldman (1986). <sup>e</sup> DePaulo, Zuckerman, and Rosenthal (1980; excluding studies overlapping with the present meta-analyses). <sup>f</sup> Zuckerman, DePaulo, and Rosenthal (1981; excluding studies overlapping with the present meta-analyses). <sup>g</sup> The median is based on five independent results. We computed one effect size each for the VA assessment study and Feldman's (1986) meta-analysis. DePaulo et al.'s (1980) and Zuckerman, DePaulo, and Rosenthal's (1981) meta-analysis were combined because they were not completely independent.

ratings was .31; the average effect size relating judge ratings and peer ratings was .30. These effect sizes are not significantly different from the overall effect size or the effect sizes of just the clinical outcomes in the present study, as can be seen in Table 7. Thus, ratings from thin slices of behavior apparently predict certain clinical criteria as well as more complicated and lengthy methods, as was advocated by Carl Rogers and his associates.

Fourth, we compared our results with a study on college teacher effectiveness. The most commonly used criterion measure for teacher effectiveness is student evaluations, a measure with high ecological validity because it is used for promotion, hiring, and tenure decisions. In a meta-analysis of the extant literature relating teacher effectiveness to aspects of teacher personality, Feldman (1986) compared college teacher effectiveness ratings that were based on student evaluations with (a) self-report measures of personality, (b) student ratings of teachers' personality, and, (c) colleague ratings of teacher personality. He conducted 14 separate meta-analyses on 14 broad personality traits evaluated in the literature. Averaging across these traits, we found that the average correlation with teacher effectiveness was as follows: (a) For self-report measures,  $r = .07$  ( $Z = 3.05$ ), (b) for students' ratings of personality,  $r = .41$  ( $Z = 13.71$ ), and (c) for colleagues' ratings of personality,  $r = .33$  ( $Z = 8.37$ ). Contrast analyses comparing the average effect size from the present meta-analysis with these three effect sizes individually as well as in combination ( $r = .28$ ) were not significant. Ratings from thin slices of behavior are apparently as good a

predictor of teaching effectiveness as other measures. This result is quite surprising because colleagues and students have access to so much more information about the subject than judges viewing clips of behavior under 300 s in length.

Our final comparison was with two meta-analyses on the accuracy of detecting deception. In both studies, no overall effect size was reported for accuracy, but effect sizes were reported for the different behavioral channels and different combinations of channels. It was not possible for us to compute overall effect sizes from the information provided in the studies. We were, however, able to compare the effect sizes for the accuracy of detection of deception from the face, body, and speech channels for both studies (leaving out results included in our analysis) with that found in the present analysis. In the first meta-analysis, the effect size for 6 studies was .32 (DePaulo et al., 1980). In the second one, the effect size for 17 studies on the accuracy of deception detection from observations of the face, body, and speech channels was .45 (Zuckerman et al., 1981). The corresponding effect size in the present analysis was .24. Again, contrast analyses revealed no significant differences between the accuracy of detecting deception among the three results ( $Z = .12$  with DePaulo et al., 1980,  $Z = .54$  with Zuckerman et al., 1981). This result suggests that thin slices of behavior may be used to predict deception about as accurately as longer observations do.

The combined median effect size of all the studies using thick slices of behavior (using only one entry per study and combining DePaulo et al.'s, 1980, and Zuckerman, DePaulo, &

Rosenthal's, 1981, meta-analyses because they are not completely independent) was .31. This figure is very close to the effect size of .39 obtained from thin slices of behavior.

## Discussion

### *Summary of Findings*

Thin slices of behavior provide a great deal of information and permit significantly accurate predictions. The effect size of .39 for the overall accuracy of prediction from observations of less than 5 min is higher than most of the effect sizes found in social and personality psychology (Cohen, 1988). An  $r$  of .39 with the criterion, according to Rosenthal and Rubin's binomial effect-size display, means that correct classifications can be made using thin slices of behavior nearly 70% of the time, compared with about 30% of the time when no thin slices are available (Rosenthal & Rubin, 1982). Furthermore, the thinness of the slice does not seem to affect the accuracy of predictions: Judgments from under 30 s of observation were as accurate as those made from 5-min observations. Indeed, the level of accuracy did not differ significantly between 30-s, 1-, 2-, 3-, 4-, and 5-min-long observations. Moreover, the accuracy of predictions from thin slices of behavior did not differ significantly from the accuracy of predictions that were based on lengthier observations of behavior, such as those in some of the classic studies on the prediction of behavior. In fact, other studies using a number of different measures to predict aspects of personality have found effect sizes ranging from .30 to .40 (Funder & Ozer, 1983). This result contradicts the commonsense notion that more information leads to greater accuracy; the additional information might be redundant, or even counterproductive (Wilson & Schooler, 1991).

Although specific behaviors exhibited within a situation might vary considerably, it appears that some stable underlying essence is picked up by judges. The consistency of predictions that are based on thin slices of behavior indicates that the "something" in the nature of people that, according to Allport (1937), leads observers to perceive them in a certain way is communicated through their expressive behavior. Individuals might not be perceived in exactly the same manner from one observation to another because of some degree of variability in their behavior; however, raters' relative ranking of individuals seems to be fairly stable (Kenrick & Funder, 1988). The data presented in this article indicate that judgments across different thin slices of behavior are quite consistent, although some people are easier to judge than others, as has been illustrated by findings regarding the "demeanor bias" (Kraut, 1982; Riggio, Tucker, & Widaman, 1987; Zuckerman, Larrance, Hall, DeFrank, & Rosenthal, 1979), and some people are better judges than others (Rosenthal et al., 1979).

### *Tentative Explanations*

Why are judgments from thin slices of behavior so accurate? We can suggest a few tentative explanations that are not mutually exclusive. The first explanation is derived from the ecological approach to social perception. The second draws on evi-

dence regarding the kernel of truth to stereotypes and the effect of self-fulfilling prophecies. The third explanation is based on evidence regarding the disruptive effects of thinking and reasoning.

The first explanation is suggested by McArthur and Baron's (1983) ecological approach to social perception. Certain attributes such as anger, fear, or dominance might be quickly and easily recognizable because they are more essential for survival and adaptive action. On the other hand, attributes like reliability or humor may be harder to detect because they are less essential for immediate survival and adaptation to the environment and require more inferential processes to identify them. Detection of the attribute depends on the context in which the target is being observed. For example, honesty may be more easy to detect when observing salesmen than when observing teachers.

Zajonc (1980, 1984) also suggested that immediate affective reactions to stimuli precede cognitive and perceptual operations. His explanation, similar to the ecological approach, was that these reactions are hypothesized to result from a primitive neurological system that allows for quick analyses and rapid action in case of threats to the organism. The initial affective reactions are based on qualities that automatically draw attention and can be evaluated at a preconscious level for favorability-unfavorability. In judging personal attributes, it is likely that attributes relating to affect, emanating from expressive behavior, allow for quick processing along a pleasant-unpleasant or a safety-threat dimension. Evidence for the preattentive processing of angry faces as opposed to happy faces in a crowd of dissimilar faces, presumably because of a preattentive search for threat, suggests this hypothesis might be true (Hansen & Hansen, 1988). Recent research on the automaticity of social information processing indicates that we do react in an automatic, affective, and evaluative manner to social stimuli (Isen, 1984) and spontaneously and automatically categorize social information into traits (Smith & Miller, 1983; Srull & Wyer, 1979; Winter & Uleman, 1984). This process seems to be automatic, in that this categorization occurs even when trait stimuli are presented subliminally (Bargh, 1988; Bargh & Pietromonaco, 1982). For example, in some of these experiments, subjects who were not even aware that a trait word had been flashed outside the visual foveal field still rated a target person as possessing more of the trait than control subjects (Bargh, Bond, Lombardi, & Tota, 1986; Bargh & Pietromonaco, 1982). Other studies have suggested that this unintentional processing is more likely to occur in the case of social stimuli, that are automatically evaluated as "good" or "bad" (Fazio, Sanbonmatsu, Powell, & Kardes, 1986). Because of some kind of non-conscious "tacit knowledge" (Polanyi, 1966), judges seem to be able to rate very brief exposures of behavior fairly accurately on various affective and evaluative dimensions.

Our findings indicate clearly that certain affective, interpersonally oriented dimensions of personality can be judged quite rapidly, efficiently, and accurately. It is certainly possible that judgments of these dimensions that are based on thin slices of behavior are accurate because recognition of these dimensions may be more important for survival and adaptation to the envi-

ronment. Furthermore, these dimensions seem to be revealed through various channels of expressive behavior.

One problem with applying this explanation to the present findings is that the judgments made by subjects in the studies reviewed here were not completely spontaneous. We cannot say whether these judgments would have been made if judges had not been instructed to do so. Some studies have indicated that subjects make trait attributions only if they are primed or instructed to do so (Bassili, 1989; Bassili & Smith, 1986; Hastie & Pennington, 1989). The stimuli in all these studies, however, consisted of descriptions of behavior. When subjects have access to actual behavior, however, as in the studies reviewed in this meta-analysis, we believe that automatic evaluation does occur.

A second, somewhat related explanation is that initial judgments are influenced by the activation in memory of common stereotypes that might possess a kernel of truth (Baron & Boudreau, 1987; McArthur, 1982; Watson, 1989). Evidence for the kernel of truth in perceptions that are based on stereotypes comes from studies relating targets' physical characteristics to judgments of various attributes of their personality (Berry, 1990; Berry & Brownlow, 1989; Berry & McArthur, 1985; Brownlow & Zebrowitz, 1990; McArthur & Montepare, 1989; Raines, Hechtman, & Rosenthal, 1990). For example, Berry and McArthur (1985) found that adults with baby faces were perceived as more honest, naive, warmer, and kinder than more mature-faced adults. Likewise, baby-faced adults guilty of criminal acts were given lighter sentences (Berry & Zebrowitz-McArthur, 1988). There is also evidence for stereotyping on the basis of vocal characteristics. People are able to discriminate vocal attractiveness and attribute dispositional characteristics to people accordingly (Zuckerman & Driver, 1989; Zuckerman, Hodgins, & Miyake, 1990).

Furthermore, evidence linking biological, physical, and temperamental attributes is increasing. Findings indicate that stable differences in the presence of certain traits in infants and children—specifically traits associated with behavioral inhibition—are potentially related to differences in underlying biological factors, especially in processes that originate in the limbic system (Kagan, Reznick, Clarke, Snidman, & Garcia-Coll, 1984; Kagan, Reznick, & Snidman, 1988; Reznick et al., 1986). Furthermore, these differences have also been associated with differences in physical characteristics in adults. Thus shy, more inhibited men seem to have more lightly colored eyes and more ectomorphic physiques than more sociable men (Herbener, Kagan, & Cohen, 1989; Rosenberg & Kagan, 1987; also see Sheldon, Stevens, & Tucker, 1940). Attempts are being made to explain why these biological, physical, and temperamental characteristics might be intercorrelated (Kagan, 1989).

Activation of these physically based stereotypes probably creates expectations in others that influence the behavior of the target individuals. Research has shown that our expectations affect our behavior toward others, which in turn modifies their behavior to confirm these expectations, creating a self-fulfilling prophecy (Anderson & Bem, 1981; Curtis & Miller, 1986; Rosenthal & Jacobson, 1968; Snyder et al., 1977). It is therefore possible that through processes such as behavioral confirmation or self-verification (Snyder et al., 1977; Swann & Read, 1981), people develop a repertoire of behaviors and a style of

interacting that validate and confirm their own and others' expectations that are based on their physical characteristics. Thus, physically attractive people who are judged to possess more socially desirable personality traits may internalize these expectations and may actually become more socially skilled, likeable, and confident (Adams, 1977; Berscheid & Walster, 1974; Dion, 1986; Goldman & Lewis, 1977). Similarly, baby-faced people may behave in more naive and less dominant ways because people expect them to behave in such ways, internalizing this view of themselves and thereby validating the kernel of truth in the stereotype.

A third explanation could be that predictions that are based on thin slices may be accurate because of the absence of distracting stimuli. Research has indicated that subjects involved in face-to-face interactions with targets were less accurate in their judgments than subjects who formed impressions from videotapes of the targets (Gilbert & Krull, 1988; Toris & DePaulo, 1984). When people are involved in actual interactions, they may be distracted by factors such as the verbal component of the interaction or the demands of impression management and self-presentation. Besides distracting external stimuli, distracting internal processing might also decrease the accuracy of judgments. Too much thinking and reasoning can sometimes be disruptive of judgmental accuracy. People make better affective judgments and decisions when they introspect less and do not seek reasons to explain their feelings (Wilson, Dunn, Kraft, & Lisle, 1989; Wilson & Schooler, 1991). Judgments that are based on thin slices of behavior may be accurate precisely because they are snap judgments. Note that this explanation contradicts the assumption of the first explanation that people can screen out distracting stimuli and focus on dimensions critical to the context in which the judgment is being made.

We are not convinced that any one of these theories alone can explain our findings. Aspects of each of these seem to influence judgments that are based on thin slices of behavior. It seems likely that the thinness of the slice eliminates distracting stimuli and enables judges to focus on expressive behavior. It is also likely that judgments regarding certain dimensions are accurate because we are used to rapidly making judgments that enable survival and adaptation to the environment. Furthermore, judgments that are based on thin slices of behavior may be accurate because they activate stereotypes that are accurate because the social world operates to reinforce and maintain certain patterns of behavior in people.

### *Implications and Conclusions*

These findings have theoretical implications for the debate in personality psychology regarding the consistency of behavior. This debate has been summarized elsewhere (Kenrick & Funder, 1988; Ross & Nisbett, 1991), and we do not discuss it in detail here. In a review of the issues in this debate, Kenrick and Funder identified the circumstances under which behavior can be predicted from trait ratings. It can be predicted when (a) publicly observable dimensions are rated, (b) raters are familiar with the target, (c) multiple raters are used, (d) multiple observations of the target are made, and (e) behaviors relative to the dimensions rated are being predicted (Kenrick & Funder, 1988). The studies included in this meta-analysis met all but

two of the criteria for the accurate prediction of behavior. First, the raters were not familiar with the target, and second, in a number of the studies surveyed multiple behavioral observations were not made. These exceptions have important implications for the issue of behavioral consistency. Raters might not have to be familiar with the target, and multiple observations of behavior might not be needed if the dimensions evaluated are truly relevant to the outcome being predicted (high validity) and if for the most part there is good agreement among the raters (high reliability) for accurate predictions, even if they are based on only thin slices of behavior. The importance of carefully selecting the traits and behaviors to be judged is highlighted by these results because certain traits are only revealed in and are only relevant to certain situations (Allport, 1966; Bem & Funder, 1978; Epstein, 1979; Funder & Drobth, 1987; Kenrick & Funder, 1988; Kenrick, McCreath, Govern, King, & Bordin, 1990). For example, the low validity of unstructured interviews in predicting job performance, college success, and professional success (Hunter & Hunter, 1984) can be attributed to the inadequate sampling of truly relevant behaviors (Ross & Nisbett, 1991). Therefore, the relevance, representativeness, and ecological validity of the behavior as well as the outcome measures are important for accurate prediction. To the degree that situations overlap and individuals are consistent in their style of behavior across different situations, these predictions should be generalizable across situations (Allport, 1937; Epstein, 1979; Kenrick & Stringfield, 1980).

Related to this issue is the issue regarding the types of behaviors or dimensions that can be judged accurately by using ratings of thin slices of behavior. It would be unrealistic to suggest that brief observations can predict most clinical and social outcomes. Brief observations may be most appropriate in predicting criterion variables characterized by observability and affectivity. First, as stated by Kenrick and Funder (1988), the behaviors or traits to be judged should be observable to permit reliability in ratings. Studies on the accuracy of personality judgments, using self-reports as a criterion and peer and stranger ratings as predictors, have generally found that observable traits and behaviors are more accurately judged than less observable ones (Albright et al., 1988; Funder & Colvin, 1988; Kenrick & Funder, 1988; Kenrick & Stringfield, 1980; Korfetzky, Kohn, & Jeger, 1978; McCrae, 1982; Watson, 1989). Thus, traits such as extraversion and conscientiousness seem to be judged more accurately than traits such as emotional stability.

Second, the dimensions or traits to be judged should include a substantial *affective* or *interpersonally oriented* component, such as teachers' expectations or patient satisfaction with doctors. Although dimensions such as anxiety, dominance, shyness, or warmth might be revealed in brief observations, less interpersonal but more personal qualities such as conscientiousness, intelligence, or persistence are probably more difficult to judge in this way. These affective, observable dimensions seem to be the ones that need to be judged quickly for survival and adaptation to the environment (McArthur & Baron, 1983; Zajonc, 1980, 1984). These interpersonal variables can be assessed even when the segment of behavior to be judged does not show an interpersonal interaction but shows only one target person. Early research has also revealed that certain personality dimensions such as inhibition-impulsion,

apathy-intensity, and ascendance-submission are judged more accurately from brief motion pictures than are dimensions such as creativity, interest in ideas and theories, or a liking for contemplative observation (Estes, 1938). This finding may be because the latter dimensions were less observable, interpersonally oriented, and less revealing of affect than the former dimensions.

These findings also have several practical implications. First, researchers can save time (their own and that of their raters) and money by using thin slices of behavior to evaluate important affective variables, without sacrificing accuracy. Second, ratings of thin slices of behavior can be used to predict important criterion variables, particularly those that are interpersonally oriented. For example, these ratings might be used to identify biased teachers, assess aspects of the therapeutic process, or gauge the expectancies of various targets such as newscasters. Third, ratings of thin slices of behavior might be very useful in the selection, training, and evaluation of people who need strong interpersonal skills, such as managers, salespersons, teachers, and therapists. Fourth, the channel of communication (verbal or nonverbal) does not seem to affect the accuracy of ratings when exposures are very brief, implying that ratings can be based on any channels that can be conveniently recorded.

In addition, these results provide additional support for the accuracy of the layperson's intuitive judgments (Funder, 1987; Kenny & Albright, 1987; Swann, 1984; Wilson & Schooler, 1991). They reveal that we unknowingly encode and decode a great deal of information regarding various aspects of ourselves. Funder proposed two criteria to evaluate the accuracy of social judgments. First, do the judgments agree with each other? Second, do they predict behavior? Most of the research considered in this article meets both criteria. Overall, judges tended to agree with each other, and their ratings did indeed predict the criterion variables. Gordon Allport (1937) observed that

a brief acquaintance often does result in amazingly rich impressions, many of which are proved on further acquaintance to be correct. Such successful judgments are significant because, lacking personal information, or a telltale context of conversation, the cues are derived entirely from expressive movements—from appearance, gesture, and manner of speaking. (p. 500)

This observation was confirmed by our findings. The probabilistic expectancies we form about others from very limited information are more accurate than we would expect.

## References

- Adams, G. R. (1977). Physical attractiveness research: Toward a developmental social psychology of beauty. *Human Development, 20*, 217-239.
- Albright, L., Kenny, D. A., & Malloy, T. E. (1988). Consensus in personality judgments at zero acquaintance. *Journal of Personality and Social Psychology, 55*, 387-395.
- Allport, G. W. (1937). *Personality: A psychological interpretation*. New York: Holt.
- Allport, G. W. (1961). *Pattern and growth in personality*. New York: Holt, Rinehart & Winston.
- Allport, G. W. (1966). Traits revisited. *American Psychologist, 21*, 1-10.

- Allport, G. W., & Vernon, P. E. (1933). *Studies in expressive movement*. New York: Haffner.
- Andersen, S. M., & Bem, S. L. (1981). Sex typing and androgyny in dyadic interactions: Individual differences in responsiveness to physical attractiveness. *Journal of Personality and Social Psychology*, *41*, 74–86.
- Apple, W., Streeter, L. A., & Krauss, R. M. (1979). Effects of pitch and speech rate on personal attributions. *Journal of Personality and Social Psychology*, *37*, 715–727.
- Archer, D., & Akert, R. M. (1977). Words and everything else: Verbal and nonverbal cues in social interpretation. *Journal of Personality and Social Psychology*, *35*, 443–449.
- Argyle, M., Alkema, F., & Gilmour, R. (1971). The communication of friendly and hostile attitudes by verbal and non-verbal signals. *European Journal of Social Psychology*, *1*, 385–402.
- Babad, E., Bernieri, F., & Rosenthal, R. (1987). Nonverbal and verbal behavior of preschool, remedial, and elementary school teachers. *American Educational Research Journal*, *24*, 405–415.
- Babad, E., Bernieri, F., & Rosenthal, R. (1989b). Nonverbal communication and leakage in the behavior of biased and unbiased teachers. *Journal of Personality and Social Psychology*, *56*, 89–94.
- Babad, E., Bernieri, F., & Rosenthal, R. (1989c). When less is more informative: Diagnosing teacher expectations from brief samples of behavior. *British Journal of Educational Psychology*, *59*, 281–295.
- Bargh, J. A. (1988). Automatic information processing: Implications for communication and affect. In L. Donohew & H. E. Sypher (Eds.), *Communication, social cognition and affect* (pp. 9–32). Hillsdale, NJ: Erlbaum.
- Bargh, J. A., Bond, R. N., Lombardi, W. J., & Tota, M. E. (1986). The additive nature of chronic and temporal sources of construct accessibility. *Journal of Personality and Social Psychology*, *50*, 869–878.
- Bargh, J. A., & Pietromonaco, P. (1982). Automatic information processing and social perception: The influence of trait information presented outside of conscious awareness on impression formation. *Journal of Personality and Social Psychology*, *43*, 437–449.
- Baron, R. M., & Boudreau, L. (1987). An ecological perspective on integrating personality and social psychology. *Journal of Personality and Social Psychology*, *53*, 1222–1228.
- Bassili, J. N. (1989). Traits as action categories versus traits as person attributes in social cognition. In J. N. Bassili, (Ed.), *On-line cognition in person perception*. Hillsdale, NJ: Erlbaum.
- Bassili, J. N., & Smith, M. C. (1986). On the spontaneity of trait attribution: Converging evidence for the role of cognitive strategy. *Journal of Personality and Social Psychology*, *50*, 239–245.
- Bem, D. J., & Funder, D. C. (1978). Predicting more of the people more of the time: Assessing the personality of the situation. *Psychological Review*, *85*, 485–501.
- Berry, D. S. (1990). Vocal attractiveness and vocal babyishness: Effects on stranger, self, and friend impressions. *Journal of Nonverbal Behavior*, *14*, 141–153.
- Berry, D. S., & Brownlow, S. (1989). Were the physiognomists right? Personality correlates of facial babyishness. *Personality and Social Psychology Bulletin*, *15*, 266–279.
- Berry, D. S., & McArthur, L. Z. (1985). Perceiving character in faces: The impact of age-related craniofacial changes on social perception. *Psychological Bulletin*, *100*, 3–18.
- Berry, D. S., & Zebrowitz-McArthur, L. (1988). What's in a face? Facial maturity and the attribution of legal responsibility. *Personality and Social Psychology Bulletin*, *14*, 23–33.
- Berscheid, E., & Walster, E. (1974). Physical attractiveness. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 7, pp. 158–216). San Diego, CA: Academic Press.
- Blanck, P. D., Rosenthal, R., & Cordell, L. H. (1985). The appearance of justice: Judges' verbal and nonverbal behavior in criminal jury trials. *Stanford Law Review*, *38*, 89–164.
- Blanck, P. D., Rosenthal, R., & Vannicelli, M. (1986). Talking to and about patients: The therapist's tone of voice. In P. D. Blanck, R. Buck, & R. Rosenthal (Eds.), *Nonverbal communication in the clinical context* (pp. 99–143). University Park: Pennsylvania State University.
- Blanck, P. D., Rosenthal, R., Vannicelli, M., & Lee, T. D. (1986). Therapists' tone of voice: Descriptive, psychometric, interactional and competence analyses. *Journal of Social and Clinical Psychology*, *4*, 154–178.
- Bond, C. F., Kahler, K. N., & Paolicelli, L. M. (1985). The miscommunication of deception: An adaptive perspective. *Journal of Experimental Social Psychology*, *21*, 331–345.
- Brown, R. (1965). *Social psychology*. New York: Free Press.
- Brown, R. (1986). *Social psychology: The second edition*. New York: Free Press.
- Brownlow, S., & Zebrowitz, L. A. (1990). Facial appearance, gender, and credibility in television commercials. *Journal of Nonverbal Behavior*, *14*, 51–60.
- Bugental, D. B., Caporael, L., & Shennum, W. A. (1980). Experimentally produced child uncontrollability: Effects of potency of adult communication patterns. *Child Development*, *51*, 520–528.
- Bugental, D. E., Henker, B., & Whalen, C. K. (1976). Attributional antecedents of verbal and vocal assertiveness. *Journal of Personality and Social Psychology*, *34*, 405–411.
- Bugental, D. E., & Love, L. (1975). Nonassertive expression of parental approval and disapproval and its relationship to child disturbance. *Child Development*, *46*, 747–752.
- Bugental, D. E., Love, L. R., Kaswan, J. W., & April, C. (1971). Verbal-nonverbal conflict in parental messages to normal and disturbed children. *Journal of Abnormal Psychology*, *77*, 6–10.
- Burns, K. L., & Beier, E. G. (1973). Significance of vocal and visual channels in the decoding of emotional meaning. *Journal of Communication*, *23*, 118–130.
- Burstein, J. W., & Carkhuff, R. R. (1968). Objective therapist and client ratings of therapist-offered facilitative conditions of low to moderate functioning therapists. *Journal of Clinical Psychology*, *24*, 240–241.
- Carkhuff, R. R., & Berenson, B. G. (1967). *Beyond counseling and therapy*. New York: Holt, Rinehart & Winston.
- Chaikin, A. L., Sigler, E., & Derlega, V. J. (1974). Nonverbal mediation of teacher expectancy effects. *Journal of Personality and Social Psychology*, *30*, 144–149.
- Cheek, J. M. (1982). Aggregation, moderator variables, and the validity of personality tests: A peer-rating study. *Journal of Personality and Social Psychology*, *43*, 1254–1269.
- Christensen, D., & Rosenthal, R. (1982). Gender and nonverbal decoding skills as determinants of interpersonal expectancy effects. *Journal of Personality and Social Psychology*, *42*, 75–87.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (Rev. ed.). San Diego, CA: Academic Press.
- Curtis, R. C., & Miller, K. (1986). Believing another likes or dislikes you: Behaviors making the beliefs come true. *Journal of Personality and Social Psychology*, *51*, 284–290.
- DePaulo, B. M., Kirkendol, S. E., Tang, J., & O'Brien, T. P. (1988). The motivational impairment effect in the communication of deception: Replications and extensions. *Journal of Nonverbal Behavior*, *12*, 177–202.
- DePaulo, B. M., Lanier, K., & Davis, T. (1983). Detecting the deceit of the motivated liar. *Journal of Personality and Social Psychology*, *45*, 1096–1103.
- DePaulo, B. M., & Rosenthal, R. (1979). Telling lies. *Journal of Personality and Social Psychology*, *37*, 1713–1722.

- DePaulo, B. M., Rosenthal, R., Eisenstat, R. A., Rogers, R. L., & Finkelstein, S. (1978). Decoding discrepant nonverbal cues. *Journal of Personality and Social Psychology*, 36, 313-323.
- DePaulo, B. M., Zuckerman, M., & Rosenthal, R. (1980). Modality effects in the detection of deception. In L. Wheeler (Ed.), *Review of Personality and Social Psychology*. Beverly Hills, CA: Sage.
- Dion, K. K. (1986). Stereotyping based on physical attractiveness: Issues and conceptual perspectives. In C. P. Herman, M. P. Zanna, and E. T. Higgins (Eds.), *Physical appearance, stigma, and social behavior: The Ontario Symposium* (Vol. 3, pp. 7-21). Hillsdale, NJ: Erlbaum.
- Duncan, S., & Rosenthal, R. (1968). Vocal emphasis in experimenters' instruction reading as unintended determinants of subjects' response. *Language and Speech*, 11, 20-26.
- Ekman, P. (1988). Lying and nonverbal behavior: Theoretical issues and new findings. *Journal of Nonverbal Behavior*, 12, 163-175.
- Ekman, P., & Friesen, W. V. (1969). Nonverbal leakage and cues to deception. *Psychiatry*, 32, 88-106.
- Ekman, P., Friesen, W. V., O'Sullivan, M., & Scherer, K. (1980). Relative importance of face, body, and speech in the judgments of personality and affect. *Journal of Personality and Social Psychology*, 38, 270-277.
- English, R. W., & Jelenevsky, S. (1971). Counselor behavior as judged under audio, visual, and audiovisual communication conditions. *Journal of Consulting Psychology*, 18, 509-513.
- Epstein, S. (1979). The stability of behavior: On predicting most of the people most of the time. *Journal of Personality and Social Psychology*, 37, 1097-1126.
- Estes, S. G. (1938). Judging personality from expressive behavior. *Journal of Abnormal and Social Psychology*, 33, 217-236.
- Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology*, 50, 229-238.
- Feldman, K. A. (1986). The perceived instructional effectiveness of college teachers as related to their personality and attitudinal characteristics: A review and synthesis. *Research in Higher Education*, 24, 139-213.
- Friedman, H. S., DiMatteo, M. R., & Mertz, T. I. (1980). Nonverbal communication on television news: The facial expression of broadcasters during coverage of a presidential election campaign. *Personality and Social Psychology Bulletin*, 6, 427-435.
- Funder, D. C. (1987). Errors and mistakes: Evaluating the accuracy of social judgment. *Psychological Bulletin*, 101, 75-91.
- Funder, D. C., & Colvin, C. R. (1988). Friends and strangers: Acquaintanceship, agreement, and the accuracy of personality judgment. *Journal of Personality and Social Psychology*, 55, 149-158.
- Funder, D. C., & Dobroth, K. M. (1987). Differences between traits: Properties associated with interjudge agreement. *Journal of Personality and Social Psychology*, 52, 409-418.
- Funder, D. C., & Ozer, D. J. (1983). Behavior as a function of the situation. *Journal of Personality and Social Psychology*, 44, 107-112.
- Gallios, C., & Callan, V. J. (1986). Decoding emotional messages: Influence of ethnicity, sex, message type, and channel. *Journal of Personality and Social Psychology*, 51, 755-762.
- Gilbert, D. T., & Krull, D. S. (1988). Seeing less and knowing more: The benefits of perceptual ignorance. *Journal of Personality and Social Psychology*, 54, 193-202.
- Goffman, E. (1959). *The presentation of self in everyday life*. Garden City, NY: Doubleday, Anchor Books.
- Goffman, E. (1971). *Relations in public*. New York: Basic Books.
- Goffman, E. (1979). *Gender advertisements*. New York: Harper & Rowe.
- Goldman, W., & Lewis, P. (1977). Beautiful is good: Evidence that the physically attractive are more socially skillful. *Journal of Experimental Social Psychology*, 13, 125-130.
- Hall, J. A. (1984). *Nonverbal sex differences: Communication accuracy and expressive style*. Baltimore: Johns Hopkins University Press.
- Hansen, C. H., & Hansen, R. D. (1988). Finding the face in the crowd: An anger superiority effect. *Journal of Personality and Social Psychology*, 54, 917-924.
- Harris, M. J. (1984). *Meta-analytic techniques and other selected statistical procedures*. Unpublished manuscript.
- Harris, M. J., & Rosenthal, R. (1985). Mediation of interpersonal expectancy effects: 31 meta-analyses. *Psychological Bulletin*, 97, 363-386.
- Hastie, R., & Pennington, N. (1989). Notes on the distinction between memory-based versus on-line judgments. In J. N. Bassili (Ed.), *Online cognition in person perception* (pp. 1-18). Hillsdale, NJ: Erlbaum.
- Herbener, E., Kagan, J., & Cohen, M. (1989). Shyness and olfactory threshold. *Personality and Individual Differences*, 10, 1159-1163.
- Holt, R. R., & Luborsky, L. (1958). *Personality patterns of psychiatrists*. (Vol. 1). New York: Basic Books.
- Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, 96, 72-98.
- Isen, A. M. (1984). Toward understanding the role of affect in cognition. In R. Wyer & T. Srull (Eds.), *Handbook of social cognition* (pp. 179-236). Hillsdale, NJ: Erlbaum.
- Kagan, J. (1989). *Unstable ideas: Temperament, cognition, and self*. Cambridge, MA: Harvard University Press.
- Kagan, J., Reznick, J. S., Clarke, C., Snidman, N., & Garcia-Coll, C. (1984). Behavioral inhibition to the unfamiliar. *Child Development*, 55, 2212-2225.
- Kagan, J., Reznick, J. S., & Snidman, N. (1988). Biological basis of childhood shyness. *Science*, 240, 167-171.
- Kenny, D. A., & Albright, L. (1987). Accuracy in interpersonal perception: A social relations analysis. *Psychological Bulletin*, 102, 390-402.
- Kenrick, D. T., & Funder, D. C. (1988). Profiting from controversy: Lessons from the person-situation debate. *American Psychologist*, 43, 23-34.
- Kenrick, D. T., McCreath, H. E., Govern, J., King, R., & Bordin, J. (1990). Person-environment intersections: Everyday settings and common trait dimensions. *Journal of Personality and Social Psychology*, 58, 685-698.
- Kenrick, D. T., & Stringfield, D. O. (1980). Personality traits and the eye of the beholder: Crossing some traditional philosophical boundaries in the search for consistency in all of the people. *Psychological Review*, 87, 88-104.
- Koretzky, M. B., Kohn, M., & Jeger, A. M. (1978). Cross-situational consistency among problem adolescents: An application of the two-factor model. *Journal of Personality and Social Psychology*, 36, 1054-1059.
- Krauss, R., Apple, W., Morency, N., Wenzel, C., & Winton, W. (1981). Verbal, vocal, and visible factors in judgments of another's affect. *Journal of Personality and Social Psychology*, 40, 312-320.
- Kraut, R. (1980, Autumn). Humans as lie detectors. *Journal of Communication*, 209-216.
- Kraut, R. E. (1982). Social presence, facial feedback, and emotion. *Journal of Personality and Social Psychology*, 42, 853-863.
- Kruglanski, A. W. (1989). The psychology of being "right": The problem of accuracy in social perception and cognition. *Psychological Bulletin*, 106, 395-409.
- Lippa, R. (1983). Expressive behavior. In L. Wheeler (Ed.), *Review of personality and social psychology* (Vol. 4, pp. 181-205). Beverly Hills, CA: Sage.
- McArthur, L. Z. (1982). Judging a book by its cover: A cognitive analy-



- sis of the relationship between physical appearance and stereotyping. In A. Hastorf & A. Isen (Eds.), *Cognitive Social Psychology* (pp. 149–211). New York: Elsevier Science.
- McArthur, L. Z., & Baron, R. M. (1983). Toward an ecological theory of social perception. *Psychological Review*, *90*, 215–238.
- McArthur, L. Z., & Montepare, J. M. (1989). Contributions of a baby-face and a childlike voice to impressions of moving and talking faces. *Journal of Nonverbal Behavior*, *13*, 189–203.
- McCrae, R. R. (1982). Consensual validation of personality traits: Evidence from self-reports and ratings. *Journal of Personality and Social Psychology*, *43*, 293–303.
- Mehrabian, A., & Wiener, M. (1967). Decoding of inconsistent communications. *Journal of Personality and Social Psychology*, *6*, 109–114.
- Milmoe, S., Novey, M. S., Kagan, J., & Rosenthal, R. (1968). The mother's voice: Postdictor of aspects of her baby's behavior. *Proceedings of the 76th Annual Convention of the American Psychological Association*, *3*, 463–464.
- Milmoe, S., Rosenthal, R., Blane, H. T., Chafetz, M. E., & Wolf, I. (1967). The doctor's voice: Postdictor of successful referral of alcoholic patients. *Journal of Abnormal Psychology*, *72*, 78–84.
- Mintz, J., & Luborsky, L. (1971). Segments versus whole sessions: Which is the better unit for psychotherapy process research? *Journal of Abnormal Psychology*, *78*, 180–191.
- Mullen, B., Futrell, D., Stairs, D., Tice, D. M., Baumeister, R. F., Dawson, K. E., Riordan, C. A., Radloff, C. E., Goethals, G. R., Kennedy, J. G., & Rosenfield, P. (1986). Newscasters' facial expressions and voting behavior of viewers: Can a smile elect a president? *Journal of Personality and Social Psychology*, *51*, 291–295.
- Noller, P. (1985). Video primacy—A further look. *Journal of Nonverbal Behavior*, *9*, 28–47.
- O'Sullivan, M., Ekman, P., & Friesen, W. V. (1988). The effect of comparisons on detecting deceit. *Journal of Nonverbal Behavior*, *12*, 203–215.
- Passini, F. T., & Norman, W. T. (1966). A universal conception of personality structure? *Journal of Personality and Social Psychology*, *4*, 44–49.
- Pittenger, R. E., Hockett, C. F., & Danehy, J. J. (1960). *The first five minutes*. Ithaca, NY: Matineau Press.
- Polanyi, M. (1966). *The tacit dimension*. New York: Doubleday.
- Raines, R. S., Hechtman, S. B., & Rosenthal, R. (1990). Physical attractiveness of face and voice: Effects of positivity, dominance, and sex. *Journal of Applied Social Psychology*, *20*, 1558–1578.
- Reznick, J. S., Kagan, J., Snidman, N., Gersten, M., Baak, K., & Rosenberg, A. (1986). Inhibited and uninhibited children: A follow-up study. *Child Development*, *57*, 660–680.
- Riggio, R. E., & Friedman, H. S. (1983). Individual differences and cues to deception. *Journal of Personality and Social Psychology*, *45*, 899–915.
- Riggio, R. E., & Friedman, H. S. (1986). Impression formation: The role of expressive behavior. *Journal of Personality and Social Psychology*, *50*, 421–427.
- Riggio, R. E., Tucker, J., & Widaman, K. F. (1987). Verbal and nonverbal cues as mediators of deception ability. *Journal of Nonverbal Behavior*, *11*, 126–145.
- Rogers, C., Gendlin, E., Kiesler, D., & Truax, C. (Eds.). (1967). *The therapeutic relationship and its impact: A study of psychotherapy with schizophrenics*. Madison: University of Wisconsin Press.
- Rosenberg, A., & Kagan, J. (1987). Iris pigmentation and behavioral inhibition. *Developmental Psychobiology*, *20*, 377–392.
- Rosenthal, R. (1966). *Experimenter effects in behavioral research*. Englewood Cliffs, NJ: Prentice-Hall.
- Rosenthal, R. (1969). Interpersonal expectations: Effects of the experimenter's hypothesis. In R. Rosenthal & R. L. Rosnow (Eds.), *Artifact in behavioral research*. San Diego, CA: Academic Press.
- Rosenthal, R. (1984). *Meta-analytic procedures for social research*. Beverly Hills, CA: Sage.
- Rosenthal, R. (1987). *Judgment studies: Design, analysis and meta-analysis*. Cambridge, England: Cambridge University Press.
- Rosenthal, R. (1990). Replication in behavioral research. *Journal of Social Behavior and Personality*, *5*, 1–30.
- Rosenthal, R., Blanck, P. D., & Vannicelli, M. (1984). Speaking to and about patients: Predicting therapists' tone of voice. *Journal of Consulting and Clinical Psychology*, *52*, 679–686.
- Rosenthal, R., & DePaulo, B. M. (1979). Sex differences accommodation in nonverbal communication. In R. Rosenthal (Ed.), *Skill in nonverbal communication: Individual differences*. Cambridge, MA: Oelgeschlager, Gunn, & Hain.
- Rosenthal, R., Hall, J. A., DiMatteo, M. R., Rogers, P. L., & Archer, D. (1979). *Sensitivity to nonverbal communication: The PONS Test*. Baltimore: Johns Hopkins University Press.
- Rosenthal, R., & Jacobson, L. (1968). *Pygmalion in the classroom*. New York: Holt, Rinehart & Winston.
- Rosenthal, R., & Rubin, D. B. (1978). Interpersonal expectancy effects: The first 345 studies. *Behavioral and Brain Sciences*, *3*, 377–415.
- Rosenthal, R., & Rubin, D. B. (1982). A simple, general purpose display of magnitude of experimental effect. *Journal of Educational Psychology*, *74*, 166–169.
- Rosenthal, R., & Rubin, D. B. (1986). Meta-analytic procedures for combining studies with multiple effect sizes. *Psychological Bulletin*, *97*, 527–529.
- Ross, L., & Nisbett, R. E. (1991). *The person and the situation: Perspectives of social psychology*. New York: McGraw-Hill.
- Scherer, K. R., Scherer, U., Hall, J. A., & Rosenthal, R. (1977). Differential attribution of personality based on multi-channel presentation of verbal and nonverbal cues. *Psychological Research*, *39*, 221–247.
- Schneider, D. J., Hastorf, A. H., & Ellsworth, P. C. (1979). *Person perception* (2nd ed.). Reading, MA: Addison-Wesley.
- Sheldon, W. H., Stevens, S. S., & Tucker, W. B. (1940). *The varieties of human physique: An introduction to constitutional psychology*. Harper: New York.
- Smith, E. R., & Miller, F. D. (1983). Mediation among the attributional inferences and comprehension processes: Initial findings and a general method. *Journal of Personality and Social Psychology*, *44*, 492–505.
- Snyder, M., Tanke, E., & Berscheid, E. (1977). Social perception and interpersonal behavior: On the self-fulfilling nature of social stereotypes. *Journal of Personality and Social Psychology*, *35*, 656–666.
- Srull, T. K., & Wyer, R. S., Jr. (1979). The role of category accessibility in the interpretation of information about persons: Some determinants and implications. *Journal of Personality and Social Psychology*, *37*, 1660–1672.
- Stiff, J. B., Miller, G. R., Sleight, C., Mongeau, P., 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.
- Strahan, C., & Zytowski, D. G. (1976). Impact of visual, vocal, and lexical cues on judgments of counselor qualities. *Journal of Counseling Psychology*, *23*, 387–393.
- Streeter, L. A., Krauss, R. M., Geller, V., Olson, C., & Apple, W. (1977). Pitch changes during attempted deception. *Journal of Personality and Social Psychology*, *35*, 345–350.
- Strong, S. R., Taylor, R. G., Bratton, J. C., & Loper, R. G. (1971). Nonverbal behavior and perceived counselor characteristics. *Journal of Counseling Psychology*, *18*, 554–561.
- Swann, W. B. (1984). Quest for accuracy in person perception: A matter of pragmatics. *Psychological Review*, *91*, 457–477.
- Swann, W. B., & Read, S. J. (1981). Acquiring self-knowledge: The

- search for feedback that fits. *Journal of Personality and Social Psychology*, 41, 1119–1128.
- Toris, C., & DePaulo, B. M. (1984). Effects of actual deception and suspiciousness of deception on interpersonal perceptions. *Journal of Personality and Social Psychology*, 47, 1063–1073.
- Truax, C. B. (1966). Therapist empathy, warmth and genuineness and patient personality change in groups psychotherapy: A comparison between interaction unit measures, time sample measures, patient perception measures. *Journal of Clinical Psychology*, 22, 225–229.
- Truax, C. B., & Carkhuff, R. P. (1967). *Toward effective counseling and psychotherapy*. Chicago: Aldine.
- Truax, C. B., Wittmer, J., & Wargo, D. G. (1971). Effects of the therapeutic conditions of accurate empathy, non-possessive warmth, and genuineness on hospitalized mental patients during group therapy. *Journal of Clinical Psychology*, 27, 137–142.
- Vincent, J. P., Friedman, L. C., Nugent, J., & Messerly, L. (1979). Demand characteristics in observations of marital interactions. *Journal of Consulting and Clinical Psychology*, 47, 557–566.
- Watson, D. (1989). Strangers' ratings of the five robust personality factors: Evidence of a surprising convergence with self-report. *Journal of Personality and Social Psychology*, 57, 120–128.
- Waxer, P. H. (1974). Therapist training in nonverbal communication: I. Nonverbal cues for depression. *Journal of Clinical Psychology*, 30, 215–218.
- Waxer, P. (1976). Nonverbal cues for depth of depression: Set versus no set. *Journal of Consulting and Clinical Psychology*, 44, 493.
- Waxer, P. H. (1977). Nonverbal cues for anxiety: An examination of emotional leakage. *Journal of Abnormal Psychology*, 86, 306–314.
- Wiggins, J. S. (1973). *Personality and prediction: Principles of personality assessment*. Reading, MA: Addison-Wesley.
- Wilson, T. D., Dunn, D. S., Kraft, D., & Lisle, D. J. (1989). Introspection, attitude change, and attitude-behavior consistency: The disruptive effects of explaining why we feel the way we do. *Advances in Experimental Social Psychology* (Vol. 22, pp. 287–343). San Diego, CA: Academic Press.
- Wilson, D. T., & Schooler, J. W. (1991). Thinking too much: Introspection can reduce the quality of preferences and decisions. *Journal of Personality and Social Psychology*, 60, 181–192.
- Winter, L., & Uleman, J. S. (1984). When are social judgments made? Evidence for the spontaneity of trait inferences. *Journal of Personality and Social Psychology*, 47, 237–252.
- Word, C. O., Zanna, M. P., & Cooper, J. (1974). The nonverbal mediation of self-fulfilling prophecies in interracial interaction. *Journal of Experimental Social Psychology*, 10, 109–120.
- Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist*, 35, 151–175.
- Zajonc, R. B. (1984). On the primacy of affect. *American Psychologist*, 39, 117–123.
- Zuckerman, M., Amidon, M. D., Bishop, S. E., & Pomerantz, S. D. (1982). Face and tone of voice in the communication of deception. *Journal of Personality and Social Psychology*, 43, 347–357.
- 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. 2–60). San Diego, CA: Academic Press.
- Zuckerman, M., & Driver, R. E. (1989). What sounds beautiful is good: The vocal attractiveness stereotype. *Journal of Nonverbal Behavior*, 13, 67–82.
- Zuckerman, M., Hall, J., DeFrank, R. S., & Rosenthal, R. (1976). Encoding and decoding of spontaneous and posed facial expressions. *Journal of Personality and Social Psychology*, 34, 966–977.
- Zuckerman, M., Hodgins, H., & Miyake, K. (1990). The vocal attractiveness stereotype: Replication and elaboration. *Journal of Nonverbal Behavior*, 14, 97–112.
- Zuckerman, M., Larrance, D. T., Hall, J. A., DeFrank, R. S., & Rosenthal, R. (1979). Posed and spontaneous communication of emotion via facial and vocal cues. *Journal of Personality*, 47, 712–733.
- Zuckerman, M., Larrance, D. T., Spiegel, N. H., & Klorman, R. (1981). Controlling nonverbal cues: Facial expressions and tone of voice. *Journal of Experimental Social Psychology*, 17, 506–524.
- Zuckerman, M., Spiegel, N. H., DePaulo, B. M., & Rosenthal, R. (1982). Nonverbal strategies for decoding deception. *Journal of Nonverbal Behavior*, 6, 171–187.

## Appendix

### Studies Included in the Meta-Analysis

- Ambady, N., & Rosenthal, R. (1990). *Half a minute: Predicting teacher effectiveness from nonverbal classroom behavior*. Manuscript submitted for publication.
- Apple, W., & Hecht, K. (1982). Speaking emotionally: The relation between verbal and vocal communication of affect. *Journal of Personality and Social Psychology*, 42, 864–875.
- Archer, D., & Akert, R. M. (1977). Words and everything else: Verbal and nonverbal cues in social interpretation. *Journal of Personality and Social Psychology*, 35, 443–449.
- Babad, E., Bernieri, F., & Rosenthal, R. (1987). Nonverbal and verbal behavior of preschool, remedial, and elementary school teachers. *American Educational Research Journal*, 24, 405–415.
- Babad, E., Bernieri, F., & Rosenthal, R. (1989a). *The keen eye of the observer: Students as judges of teachers' verbal and nonverbal behavior*. Manuscript submitted for publication.
- Babad, E., Bernieri, F., & Rosenthal, R. (1989b). Nonverbal communication and leakage in the behavior of biased and unbiased teachers. *Journal of Personality and Social Psychology*, 56, 89–94.
- Blanck, P. D., Rosenthal, R., & Vannicelli, M. (1986). Talking to and about patients: The therapist's tone of voice. In P. D. Blanck, R. Buck, & R. Rosenthal (Eds.), *Nonverbal communication in the clinical context*. University Park: Pennsylvania State University.
- Blanck, P. D., Rosenthal, R., Vannicelli, M., & Lee, T. D. (1986). Therapists' tone of voice: Descriptive, psychometric, interactional and competence analyses. *Journal of Social and Clinical Psychology*, 4, 154–178.
- Bugental, D. B., Love, L. R., Kaswan, J. W., & April, C. (1971). Verbal-nonverbal conflict in parental messages to normal and disturbed children. *Journal of Abnormal Psychology*, 77, 6–10.
- DePaulo, B. M., Lanier, K., & Davis, T. (1983). Detecting the deceit of the motivated liar. *Journal of Personality and Social Psychology*, 45, 1096–1103.
- DePaulo, B. M., Lassiter, G. D., & Stone, J. I. (1982). Attentional determinants or success at detecting deception and truth. *Personality and Social Psychology Bulletin*, 8, 273–279.
- DePaulo, B. M., & Rosenthal, R. (1979). Telling lies. *Journal of Personality and Social Psychology*, 37, 1713–1722.
- DePaulo, B. M., Rosenthal, R., Green, C. R., & Rosenkrantz, J. (1982).

- Diagnosing deceptive and mixed messages from verbal and nonverbal cues. *Journal of Experimental Social Psychology*, 18, 433-446.
- DePaulo, B. M., Stone, J. I., & Lassiter, G. D. (1985). Telling ingratiating lies: Effects of target sex and target attractiveness on verbal and nonverbal deceptive success. *Journal of Personality and Social Psychology*, 48, 1191-1203.
- Feldman, R. S. (1976). Nonverbal disclosure of teacher deception and interpersonal affect. *Journal of Educational Psychology*, 68, 807-816.
- Feldman, R. S., Jenkins, L., & Popoola, O. (1979). Detection of deception in adults and children via facial expressions. *Child Development*, 50, 350-355.
- Feldman, R. S., & Prohaska, T. (1979). The student as Pygmalion: Effect of student expectation on the teacher. *Journal of Educational Psychology*, 71, 485-493.
- Friedman, H. S., Hall, J. A., & Harris, M. J. (1985). Type A behavior, nonverbal expressive style, and health. *Journal of Personality and Social Psychology*, 48, 1299-1315.
- Fugita, S. S., Hoguebe, M. C., & Wexley, K. N. (1980). Perceptions of deception: Perceived expertise in detecting deception, successfulness of deception and nonverbal cues. *Personality and Social Psychology Bulletin*, 6, 637-643.
- Hall, J. A., & Braunwald, K. G. (1981). Gender cues in conversations. *Journal of Personality and Social Psychology*, 40, 99-110.
- Hall, J. A., Roter, D. L., & Katz, N. R. (1987). Task versus socioemotional behaviors in physicians. *Medical Care*, 25, 399-412.
- Hall, J. A., Roter, D. L., & Rand, C. S. (1981). Communication of affect between patient and physician. *Journal of Health and Social Behavior*, 22, 18-30.
- Kaul, T. J., & Schmidt, L. D. (1971). Dimensions of interviewer trustworthiness. *Journal of Counseling Psychology*, 18, 542-548.
- Machida, S. (1986). Teacher accuracy in decoding nonverbal indicants of comprehension and noncomprehension in Anglo- and Mexican-American children. *Journal of Educational Psychology*, 78, 454-464.
- Manstead, A. S. R., Wagner, H. L., & MacDonald, C. J. (1986). Deceptive and nondeceptive communications: Sending experience, modality, and individual abilities. *Journal of Nonverbal Behavior*, 10, 147-167.
- Milmoe, S., Novey, M. S., Kagan, J., & Rosenthal, R. (1968). The mother's voice: Postdictor of aspects of her baby's behavior. *Proceedings of the 76th Annual Convention of the American Psychological Association*, 3, pp. 463-464.
- Milmoe, S., Rosenthal, R., Blane, H. T., Chafetz, M. E., & Wolf, I. (1967). The doctor's voice: Postdictor of successful referral of alcoholic patients. *Journal of Abnormal Psychology*, 72, 78-84.
- Mullen, B., Futrell, D., Stairs, D., Tice, D. M., Baumeister, R. F., Dawson, K. E., Riordan, C. A., Radloff, C. E., Goethals, G. R., Kennedy, J. G., & Rosenfeld, P. (1986). Newscasters' facial expressions and voting behavior of viewers: Can a smile elect a president? *Journal of Personality and Social Psychology*, 51, 291-295.
- O'Sullivan, M., Ekman, P., & Friesen, W. V. (1988). The effect of comparison on detecting deceit. *Journal of Nonverbal Behavior*, 12, 203-215.
- Riggio, R. E., & Friedman, H. S. (1983). Individual differences and cues to deception. *Journal of Personality and Social Psychology*, 45, 899-915.
- Riggio, R. E., Tucker, J., & Throckmorton, B. (1987). Social skills and deception ability. *Personality and Social Psychology Bulletin*, 13, 568-577.
- Rosenthal, R., Blanck, P. D., & Vannicelli, M. (1984). Speaking to and about patients: Predicting therapists' tone of voice. *Journal of Consulting and Clinical Psychology*, 52, 679-686.
- Roter, D. L., Hall, J. A., & Katz, N. R. (1987). Relations between physicians' behaviors and analogue patients' satisfaction, recall, and impressions. *Medical Care*, 25, 437-451.
- Steckler, N. A., & Rosenthal, R. (1985). Sex differences in nonverbal and verbal communication with bosses, peers, and subordinates. *Journal of Applied Psychology*, 70, 157-163.
- Stiff, J. B., Miller, G. R., Sleight, C., Mongeau, P., 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.
- Streeter, L. A., Krauss, R. M., Geller, V., Olson, C., & Apple, W. (1977). Pitch changes during attempted deception. *Journal of Personality and Social Psychology*, 35, 345-350.
- Waxer, P. H. (1974). Therapist training in nonverbal communication: I. Nonverbal cues for depression. *Journal of Clinical Psychology*, 30, 215-218.
- Waxer, P. (1976). Nonverbal cues for depth of depression: Set versus no set. *Journal of Consulting and Clinical Psychology*, 44, 493.
- Waxer, P. H. (1977). Nonverbal cues for anxiety: An examination of emotional leakage. *Journal of Abnormal Psychology*, 86, 306-314.
- Zuckerman, M., DeFrank, R. S., Hall, J. A., Larrance, D. T., & Rosenthal, R. (1979). Facial and vocal cues of deception and honesty. *Journal of Experimental Social Psychology*, 15, 378-396.
- Zuckerman, M., Driver, R., & Guadagno, N. S. (1985). Effects of segmentation patterns on the perception of deception. *Journal of Nonverbal Behavior*, 9, 160-168.
- Zuckerman, M., Fischer, S. A., Osmun, R. W., Winkler, B. A., & Wolfson, L. R. (1987). Anchoring in lie detection revisited. *Journal of Nonverbal Behavior*, 11, 4-12.
- Zuckerman, M., Koestner, R., & Alton, A. O. (1984). Learning to detect deception. *Journal of Personality and Social Psychology*, 46, 519-528.
- 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.

Received October 29, 1990

Revision received June 28, 1991

Accepted July 1, 1991 ■