Psychological Bulletin
1987, Vol. 102, No. 3, 390–402

0033-2909/87/$00.75

Accuracy in Interpersonal Perception: A Social Relations Analysis

David A. Kenny and Linda Albright
University of Connecticut

For the past 30 years, the study of accuracy in person perception has been a neglected topic in social and personality psychology. Research in this area was stopped by a critique of global accuracy scores by Cronbach and Gage. They argued that accuracy should be measured in terms of components. Currently, interest in the topic of accuracy is rekindling. This interest is motivated, in part, by a reaction to the bias literature. We argue that modern accuracy research should (a) focus on measuring when and how people are accurate and not on who is accurate, (b) use each person as both judge and target, and (c) partition accuracy into components. The social relations model (Kenny & La Voie, 1984) can be used as a paradigm to meet these requirements. According to this model, there are four types of accuracy, only two of which are generally conceptually interesting. The first, called *individual accuracy,* measures the degree to which people's judgments of an individual correspond to how that individual tends to behave across interaction partners. The second, called *dyadic accuracy,* measures the degree to which people can uniquely judge how a specific individual will behave with them. We present an example that shows high levels of individual accuracy and lower levels of dyadic accuracy.

The topic of accuracy in interpersonal perception is a fundamental issue in social and personality psychology. In this article, we present a new and integrative approach. We begin with a historical review of the topic and the Cronbach and Gage critique of global accuracy scores. We then propose that accuracy research should be nomothetic, interpersonal, and componential. Finally, we show how the social relations model fulfills these requirements and so provides a methodology to study interpersonal accuracy.

Historical Survey

Accuracy in person perception is one of the oldest topics in social and personality psychology. The roots of this research lie in the success of standardized intelligence testing. Researchers reasoned that if it was possible to measure individual differences in cognitive skills, it should be possible to measure individual differences in social skills. Psychologists rushed to the task of measuring individual differences in accuracy in person perception. Whether it was called *accuracy, empathy, social skills, understanding,* or *sensitivity,* the goal was always the same: to differentiate people in their ability to know the social world surrounding them.

The individual-difference orientation fostered during World War II easily absorbed this tradition. American social scientists were eager to select men who could be leaders and be responsive to the demands of the men that they commanded.

After World War II, the aims continued to focus on selection. But now the emphasis was in the selection of clinicians, social workers, and teachers, who were thought to be skilled perceivers of people. Also, poorly adjusted people were thought to be those who were not accurate perceivers. In the late 1940s and early 1950s, the study of individual differences in accuracy became a dominant area of research in social and personality psychology.

Critique of Accuracy Research

It all came to a crashing halt. A number of prominent researchers, most notably Cronbach and Gage (Cronbach, 1955, 1958; Gage & Cronbach, 1955; Gage, Leavitt, & Stone, 1956), called into question the measurement techniques of the accuracy researchers. These researchers did not argue that accuracy could not be measured, as is sometimes mistakenly thought, but that a complete treatment of accuracy required much more complicated procedures than those available at that time. Because these criticisms are so important and not well understood, we review them in detail. (For even more detail, consult Kenny, 1986.) We show that the approach taken in this article parallels the Cronbach (1955) components.

Cronbach (1955) distinguished among four components of accuracy. They are elevation, differential elevation, stereotype accuracy, and differential accuracy. To understand these terms, consider the judgment process.

Each judge rates a set of targets on a set of traits. For each judgment, there is a criterion score. Accuracy is defined as the correspondence between the judgment and the criterion. Cronbach (1955) criticized the use of a single global discrepancy score as a measure of accuracy. Such a measure is the average of the discrepancies between the judgments and the criterion.

This research was supported in part by Grant R01MH4029501 from the National Institute of Mental Health.

We thank Ronald Anderson, who generously provided us with his data. Also, Bella M. DePaulo, Larry Galpert, Judith Hall, Doug Kenrick, Brewster Smith, and William Swann provided us with valuable comments on an earlier draft. Finally, Raymond R. Reno assisted us in the data analysis.

Correspondence concerning this article should be addressed to David A. Kenny, Department of Psychology, Box U-20, University of Connecticut, Storrs, Connecticut 06268.

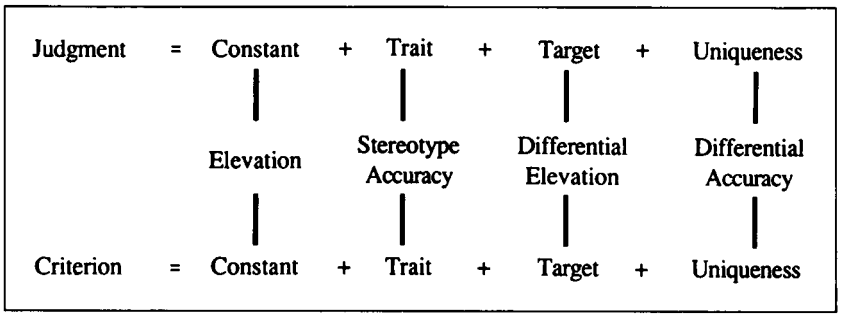

Figure 1. Cronbach's (1955) four components of accuracy.

According to Cronbach[1] (1955), the judgment and the criterion scores can each be divided into component parts. In equation form, this partitioning for the judgment is as follows: Judgment = Constant + Trait + Target + Uniqueness. The constant term is the tendency for the judge to rate all the targets on all the traits in the same direction. Ordinarily, higher scores indicate a more favorable response, so the constant indicates the extent to which the judge tends to rate others favorably or unfavorably. The constant represents a very general response set, because it affects the judge's rating of every target on every trait. The trait effect represents the judge's tendency to view a particular trait as being high or low across targets relative to the other traits that are rated. It measures the differential response set for the set of traits. So both the constant and trait terms are measures of the judge's response set.

The target effect represents any tendency for the judge to view one target more or less favorably across the set of traits than the other targets. The final component of the judgment score is what is called *uniqueness,* which represents the judge's view of the target on a particular trait after the constant, the trait effect, and the target effect are removed. Basically, this measure indicates how the target is uniquely evaluated on the particular trait by the judge.

The partitioning of the judgment represents a two-way analysis of variance (ANOVA) in which the rows are traits and the columns are targets. What we have called *uniqueness* corresponds to the target-by-trait interaction in the two-way ANOVA.

The criterion measure can be divided into the same components as the judgments. In equation form, the partitioning is as follows: Criterion = Constant + Trait + Target + Uniqueness. For instance, the trait effect measures the extent to which the traits differ on the criterion.

We are now in a position to discuss Cronbach's (1955) four components. The four components linking the judgment and criterion are diagrammed in Figure 1. In the top of the figure, the judgment is divided into four parts—constant, trait, target, and uniqueness. In the bottom, the criterion score is also divided into constant, trait, target, and uniqueness. Cronbach's four components of accuracy can be viewed as linking the corresponding parts of the judgment and criterion scores.

Elevation concerns the degree of correspondence between the constant of the judgment and the constant of the criterion; that is, it deals with the discrepancy between the judge's average score (across targets and traits) and the average score across targets and traits on the criterion. Stereotype accuracy concerns the degree of correlation between the trait effects of the judgment and the trait effects of the criterion. This component of accuracy concerns whether the pattern of the average ratings of the traits (across targets) of a judge corresponds to the pattern of the average score for the traits on the criterion. Differential elevation concerns the degree of correlation between the target effects of the judgments with the target effects of the criterion. This component of accuracy concerns whether the pattern of the average ratings of the targets of a judge corresponds to the pattern of the average score for the targets on the criterion. Differential accuracy concerns the correspondence between corresponding uniqueness components. One way to measure differential accuracy is to correlate the uniqueness scores of the judgments with the uniqueness scores of the criterion.[2]

According to Cronbach (1955) and others, only two of the four components that we have defined reflect true accuracy: differential elevation and differential accuracy. The remaining components—elevation and stereotype accuracy[3]—involve the match between the judge's response set and the criterion. Without a decomposition of accuracy into components, one would not know whether the accuracy reflects "true" or "false" accuracy (e.g., stereotype accuracy).

Postcritique Accuracy Research

Although surely unintended, accuracy research received a stigma. It became an unresearchable topic. No one wanted to

[1] In 1958, Cronbach suggested an approach entirely different from the one in his 1955 article. As we do in the approach that we suggest, he emphasized analyzing trait by trait instead of computing accuracy across the set of traits. So our expression of a "Cronbach analysis" refers to the 1955 article and not to his views since 1958.

[2] For all but elevation, Cronbach (1955) discussed two aspects of accuracy. We have already discussed the first, the correlation between the judgment component and the criterion component. The second concerns the reduction in variance of the judgments given the degree of correlation between the judgment and the criterion. This variance reduction is especially important when the judges know the unit of measurement of the criterion. Because this is often not the case, we concentrate on the correlational measure of accuracy and not the variance reduction measure.

[3] There are circumstances in which stereotype accuracy should be considered as "true" accuracy. It may reflect sensitivity to the generalized other (Bronfenbrenner, Harding, & Gallwey, 1985) and not response-set similarity.

investigate an area that was, according to Cline (1964), a "Pandora's box of 'components, artifacts, and methodological problems' (Cronbach, 1955)" (p. 227). A few brave souls continued to work on the topic, but it is fair to say that accuracy as an area of study withered away and students were advised that it was a dead topic. Cook (1979) characterized the reaction as follows:

> The whole business of trying to measure the accuracy of person perception is so hopelessly complicated that it should be abandoned. This was the impression created on many researchers by Cronbach's critiques; the apparent difficulty of doing research led many workers in the field, by a familiar rationalization, to argue that the issue wasn't important, wasn't worth studying experimentally, or even that it didn't exist. (p. 118)

Accuracy research "lost . . . its charm" (Schneider, Hastorf, & Ellsworth, 1979, p. 222).

What was to replace accuracy as an area of research? The criticism extended beyond accuracy to any measure that was dyadic. Clearly, individual topics were safer and less subject to the rapier criticisms of such methodological experts as Cronbach. The field turned to attitudes in general and dissonance theory in particular. The current fascination with intrapsychic, cognitive topics in social psychology is due, in part, to the Cronbach-Gage critique.

Research in person perception continued. Gage and Cronbach (1955) correctly predicted the dominant theme of research in person perception:

> Social perception as measured is a process dominated far more by what the Judge brings to it than by what he takes in during it. His favorability toward the Other, before or after he observes the Other, and his implicit personality theory, formed by his experiences prior to his interaction with the Other, seem to determine his perceptions. (p. 420)

No doubt, too, the "new look in perception" encouraged the field of person perception to move away from the study of accuracy and to study bias. Subsequent work in person perception that carefully documents the human observer's use of heuristics, implicit assumptions, and egocentric orientation got its impetus from the end of accuracy research.

Even if one were willing to do Cronbach (1955) analyses, the computational burden in that precomputer era was excessive. Most researchers already viewed the pre-Cronbach-and-Gage procedures available at the time as complicated enough. The suggested added complexity was too much. Various computations could not be done "because the amount of calculation involved in obtaining them is prohibitive" (Cline & Richards, 1960, p. 5). The results of all these computations were very disappointing, and Cook (1979) drew the conclusion that using "more refined methods show that perceptions of other people are for the most part very inaccurate" (p. 145).

Resurgence of Accuracy Research

The extensive bias literature (cf. Higgins, Herman, & Zanna, 1981; Nisbett & Ross, 1980) and the paucity of accuracy research has given us a potentially misleading picture of the person perceiver (Funder, 1987). We know that observers make errors, but mistakes mean only that person perceivers are not perfect. An expert tennis player double faults, makes unforced errors, and allows his or her opponent to make passing shots. Very good batters in baseball are out two thirds of the time. Excellence and perfection are not synonymous. Hastie and Rasinski (in press) showed that even though human observers make mistakes, their accuracy can be quite high. Most tests of bias take as the null hypothesis that people are totally accurate and show, not surprisingly, that indeed they are not perfectly accurate. To determine the level of accuracy, one must directly measure it and not infer it from a measure of bias.

Others besides Hastie and Rasinski (in press) have argued that person perceivers may be more accurate than one might think. McArthur and Baron (1983) took an ecological approach. In part, they argued that an experimental context with verbal stimuli is not representative of the typical human judgment situation. Kruglanski and Ajzen (1983) argued that the process of human inference has been confused with the specific instances of inference. Without a criterion measure, accuracy cannot be assessed. Swann (1984) stated that accuracy lies not in judging people in general but in judging specific interaction partners.

Second Wave of Accuracy Research

If investigators want to see a rebirth of accuracy research, they must be careful to realize the complexities raised in the 1950s. In the hurry to study accuracy, they must not repeat the mistakes of the past. Therefore, modern work on accuracy, the second wave, must be responsive to critiques of research from the 1950s. We believe that accuracy research must be nomothetic, interpersonal, and componential.

Nomothetic Orientation

Most of the research in accuracy has been in the area of individual differences. Recall that the initial impetus for research in accuracy was from intelligence testing. Both during World War II and in the postwar ear, accuracy research had an avowed selection purpose: to select either the very able or the very unable.

A number of converging sources of evidence point to small amounts of individual differences in accuracy. First, researchers who followed Cronbach and Gage measured the reliability of differential accuracy. Their general finding is that reliability in this component is low. Cronbach (1955) reported the reliability of differential accuracy as .18. Crow and Hammond (1957, Study 2) obtained overtime reliabilities in the .25 range for their measure of differential accuracy. Also, Bronfenbrenner, Harding, and Gallwey (1958) found reliability for this component to be nil (p. 52). These low reliabilities are not just true for older studies. In a more recent study, Anderson (1984) found an average reliability of .18 for the differential accuracy of four different traits. Cronbach's doubts in 1955 about "whether accuracy in differentiating personalities of others can be reliably measured" (p. 185) appear to be borne out.[4]

[4] The study by Cline and Richards (1960) is frequently cited as a post-Cronbach study showing the generalizability of individual differences. In a subsequent article, however, Richards and Cline (1963) noted that they had made a mistake in their measure of differential accuracy. The proper measures of differential accuracy show modest correlations with other measures of accuracy.

These low reliabilities can give a mistaken notion about validity. Measures of reliability assess the consistency of individual differences. If the reliability is low, it does not necessarily indicate that the average level of response is meaningless. A test can have no reliability, yet the mean can be interpretable. Imagine the following test of visual acuity in a classroom. An instructor writes a word on the blackboard and asks the students to copy it. The instructor does this 10 times. One can create a score from 0 to 10 to measure acuity. Presumably, most students would get 10s, but for various reasons, there may be a few scores of 9. The researcher computes the mean and finds that the average score for the class is 9.85. The instructor concludes that the class can read what is written on the board. Then, as an afterthought, the instructor computes a measure of reliability. Shockingly, the reliability is .04. Is the test a reliable measure? Yes and no. No, it is a poor measure of individual differences. Yes, it can determine whether the class can read what is on the board. People can be highly accurate, but the test can be totally unreliable.

This confusion of the reliability of individual differences and the reliability of accuracy scores is nowhere more evident than in Crow and Hammond's (1957) Study 1. They developed 15 measures of accuracy. As they emphasized, these measures do not intercorrelate, which casts doubt on the reliability of the measures. Of the 12 measures for which it is possible to determine whether the subjects did better than chance, however, the subjects scored significantly above chance on 11. (The remaining measure showed significant performance below chance.) Their data show remarkable levels of accuracy in the face of low consistency.

A second source of evidence of the limited individual differences in accuracy is in the area of nonverbal sensitivity. Individual differences in this area have proved to be elusive. The reliable measures Communication of Affect Receiving Ability Test (CARAT; Buck, 1984) and Profile of Nonverbal Sensitivity (Rosenthal, Hall, DiMatteo, Rogers, & Archer, 1979) appear not to correlate with each other. Kenny and La Voie's (1984) analysis indicated that individual differences in receiving and decoding ability are small. These analyses are independently supported by Bond, Kahler, and Paolicelli's (1985) data that show individual differences in lie and truth detection to be modest. Also, attempts to improve people's skills in this area have not been very successful (Zuckerman, Koestner, & Alton, 1984), a fact consistent with the view of minimal individual differences.

The final evidence concerns the studies that preceded the Cronbach and Gage critique. Certainly, a major reason that these two researchers became interested in the accuracy issue is that studies showing individual differences in accuracy failed to replicate. This failure may not have been due so much to methodological shortcomings but rather to insufficient variance.

Therefore, in this article, we downplay issues in the area of individual differences. Instead, the focus is nomothetic. The focus is not on who is accurate but on when and how people are accurate. Our position is not that individual differences are nonexistent in interpersonal accuracy. Rather, we believe that the variability of such differences is rather limited and that studying the level of accuracy is likely to be more productive.

Interpersonal Orientation

Person perceivers in everyday life do not view their targets through one-way mirrors. They touch, yell at, and interact with each other. In a related vein, it is totally arbitrary to label one of the participants the judge and the other the target because both people are judging each other (Tagiuri, 1969). Social perception is a two-sided experience. In a review of accuracy studies, Smith (1966, p. 26) noted that 56% of the studies involve judgments of targets with whom the judge has interacted. So social interaction is the rule, not the exception.

Swann (1984) carefully noted the role that interaction can play in enhancing interpersonal accuracy. He criticized the dominant use of object-perception models in the area of person perception. One problem with object-perception models is the assumption that the stimulus does not change when it is perceived by different perceivers. In person perception, the stimulus can change when it interacts with different perceivers.

Our argument that accuracy research should be interpersonal is not equivalent to Funder's (1987) argument that accuracy should be studied only in the real world. The issue is not where accuracy is studied but rather what type of stimuli should be used to assess accuracy (real people with whom one can interact versus verbal descriptions of people). The type of stimuli used in typical laboratory research in person perception can be found in the real world; one often makes judgments of individuals with whom one has not interacted. Thus, the mistakes or errors people make in the laboratory would probably be made in the real world if the context in the real world were similar to that in the laboratory. But, to assess the accuracy of interpersonal perception, one should use an interactive context. This does not, however, preclude accuracy research in the laboratory. Interaction can occur in a laboratory context (see, e.g., DePaulo, Kenny, Hoover, Webb, & Oliver, 1987).

An interactive context is also important for methodological reasons. Adopting the terms of Tagiuri, Bruner, and Blake (1958), there are three aspects to social perception. First, there is mutuality or reciprocity. If A likes B, does B like A? Second, there is congruence or assumed reciprocity. If A likes B, does A think that B likes A? And third, there is accuracy. If B likes A, does A think that B likes A? These three aspects can be viewed as forming a triangle, as in Figure 2. These three aspects are not independent. If A likes B and B likes A (mutuality) and A then assumes that B likes A (congruence), then A must be accurate at knowing that B likes A. So accuracy can be a by-product of mutuality and congruence. This potential confound can be measured and controlled only by studying both people in the dyad. The second wave of accuracy research must allow for and take into account the two-sided nature of social perception.

Componential Orientation

The essence of the Cronbach and Gage critique is that judgments must not be viewed globally but must be broken down into components. Accuracy is then measured by the correspondence between these components. Some of these components tap the subjects' response set, and so correspondence between these components does not measure "true" accuracy.

Although they often pay lip service to the Cronbach and Gage

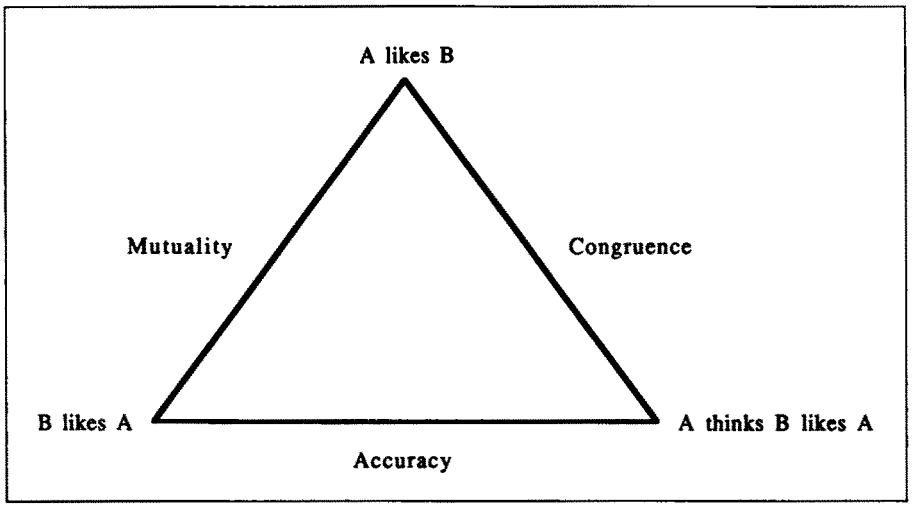

Figure 2. Mutuality, congruence, and accuracy triangle.

critique in their introductions, most contemporary researchers compute global measures of accuracy in their Results sections. Because the Cronbach and Gage critiques occurred a generation ago, many contemporary accuracy researchers are unaware of the difficulties. Although there are notable exceptions (Harackiewicz & DePaulo, 1982), contemporary accuracy research is not much better in methodology than pre-1955 research. Ironically, some pre-Cronbach articles, for example, Ausubel, Schiff, and Gasser's (1952), contain more sophisticated analyses than does a good deal of contemporary work. Modern accuracy researchers must seriously confront the Cronbach-Gage critique. Because researchers today have easy access to high-speed computers, the computational obstacles confronted by early researchers are no longer present.

Social Relations Model

If the second wave of accuracy research is to be nomothetic, interpersonal, and componential, it will need a new methodology. The social relations model can be applied to the study of accuracy and be that new methodology (Kenny, 1981; Kenny & La Voie, 1984; Malloy & Kenny, 1986).

In the Cronbach (1955) partitioning, the target-by-trait matrix is partitioned for each judge. Because our focus is nomothetic, the partitioning that occurs is of the judge-by-target matrix for each trait. That is, the classical approach is to measure the accuracy for each judge across the set of targets and traits; our approach is to measure the accuracy for a given trait across the set of judges and targets.

Because our focus is on the interpersonal nature of accuracy, we allow for the possibility that a person serves as both judge and target. Also, we allow the criterion to be different for each judge. In prototypical accuracy research, the criterion does not change. That is, all judges' responses are compared with the same criterion score. In the next section, we consider a criterion of this type.

Imagine two acquainted persons, Al and Bob. They are asked to judge how competitive each other are. Al and Bob then interact in a structured situation (e.g., the Prisoner's Dilemma Game), and both Al and Bob's competitiveness are measured. These measurements are the criterion scores. Consider now the question of how accurate Al is at judging how competitive Bob is.

In the social relations model, Al's judgment of Bob's competitiveness when interacting with Al is assumed to equal the following:

$$\text{Al's Judgment of Bob} = \text{Constant} + \text{Al's Actor Effect} + \text{Bob's Partner Effect} + \text{Al's Relationship Effect With Bob}$$

The terms of the equation are elaborated in Table 1.

The constant represents the average judgment of competitiveness across the set of judges and targets. The actor effect represents the average tendency for Al to believe that others are competitive. The partner effect represents the tendency for judges to believe that Bob is competitive. The relationship effect measures the tendency for Al to believe that Bob is particularly competitive or cooperative when interacting with Al.

The criterion score (how competitive Bob is when interacting with Al) can be partitioned into constant, actor, partner, and relationship. Its equation is as follows:

$$\text{Bob's Behavior With Al} = \text{Constant} + \text{Bob's Actor Effect} + \text{Al's Partner Effect} + \text{Bob's Relationship Effect With Al}$$

These components are also explained in Table 1. The constant represents the tendency for targets to behave competitively or cooperatively with their interaction partners. This term does not vary across judges or targets. The actor effect represents the tendency for Bob to behave competitively or cooperatively across all his interaction partners. The partner effect represents the extent to which Al's interaction partners generally behave

Table 1
Components of the Social Relations Model

Component	Al's rating of Bob's competitiveness (judgment)	Bob's competitiveness with Al (criterion)
Constant	People's general rating of others' competitiveness	People's general level of competitiveness
Actor	Al's general prediction of other's competitiveness	Bob's general level of competitiveness
Partner	Others' general rating of Bob's competitiveness	Others' general level of competitiveness when interacting with Al
Relationship	Al's rating of Bob's competitiveness controlling for Al's actor effect and Bob's partner effect	Bob's level of competitiveness with Al controlling for Bob's actor effect and Al's partner effect

competitively or cooperatively. And the relationship effect is the unique level of competitiveness in Bob when he interacts with Al. The relationship effect is directional. Bob's unique level of competitiveness with Al may not match Al's unique level with Bob.

As in Figure 1, for the Cronbach components, accuracy in the social relations model can be conceptualized as the linking together of components; this is illustrated in Figure 3. As in the Cronbach analysis, there are four types of nomothetic accuracy.

Elevation accuracy concerns the match between the judges' average response and the average response on the criterion measure. It is measured by the difference in the overall means (across judges and targets) between the judgment and the criterion. It would be reasonable to measure only if the two variables were expressed in the same unit of measurement.

Response-set accuracy concerns whether the judge's average response (actor effect on the judgment) corresponds to the average score of his or her interaction partners (his or her partner effect on the criterion). Because a judge's average score will often largely reflect a response set, we call this type of accuracy *response-set accuracy.* As with stereotype accuracy (see Footnote 3), response-set accuracy may in some cases be informative. For instance, in the competitiveness example, people who expect their partners to be competitive may create that competitiveness.

Individual accuracy concerns the extent to which a person's behavior across interaction partners corresponds to how the individual is generally viewed. It measures the correlation between the partner effect in the judgments and the actor effect in the criterion. In other words, it measures the correlation between how one is generally predicted to behave (the partner effect in judgment) and how one actually behaves across interaction partners (the actor effect in the criterion).

Dyadic accuracy concerns the ability of a judge to predict her or his partner's behavior over and above the ability of other judges' predictions of the partner's behavior. It measures the correspondence between relationship components. For instance, in the competitiveness example, it measures a judge's ability to uniquely predict a given partner's behavior with the judge. Can people differentially judge how individuals differentially behave with them?

The terms *individual* and *dyadic accuracy* refer to the level of analysis and not to the specific content of the judgments. Individual accuracy measures correspondence between others' judgments of one person and the behavior of that person. Thus, the analysis is at the individual level. Dyadic accuracy, on the other hand, measures the correspondence between differential judgments made by the judge and the differential behavior of a specific partner, and so it is at the dyadic level of analysis. These terms apply even in the study of other accuracy phenomena. For instance, in the example we describe later, the accuracy being studied is not accuracy in judging others but accuracy in knowing how one is judged by others.[5]

The social relations model is a random-effects model. That is, judges and targets are viewed as a random sample of people, and so results can be generalized beyond the particular people sampled. Because the model is random effects, estimation of individual and dyadic accuracy is complicated. Details were presented by Warner, Kenny, and Stoto (1979), and Kenny (1981). Interpretation issues were presented by Malloy and Kenny (1986).

Our measures of accuracy are related to but are not identical to other formulations of accuracy. In a sense, individual accuracy corresponds to differential elevation, and dyadic accuracy corresponds to differential accuracy in the Cronbach system. But this is only roughly the case because we are looking at the accuracy for a trait across a set of judges and targets, whereas Cronbach examined the accuracy of a judge across a set of targets and traits (Kenny, 1981).

Swann (1984), as well as McHenry (1971), defined two types of accuracy. Swann defined global accuracy as the ability to predict how a person behaves in general with others and circumscribed accuracy as the ability to predict the behavior of a person when in the presence of the judge. His global accuracy corresponds to our individual accuracy, and his circumscribed accuracy corresponds to our dyadic accuracy. In Swann's system, however, global and circumscribed accuracy are confounded. If people know how someone behaves in general, they must know, at least in part, how the person behaves when interacting with them. Such a confound is not present in individual and dyadic accuracy because in dyadic accuracy, the actor and partner effects have been removed and the correlation is purely dyadic.

Before examining the accuracy correlations, one must first consider the variance partitioning. If there is no variance on the components of interest, it makes no sense to look at the correlations. For example, for people to be individually accurate, there must be actor variance on the criterion. That is, behavior must be consistent across interaction partners.

The social relations model's four types of accuracy are nomo-

[5] If accuracy of impressions is assessed, it is more appropriate to call the correlation of the actor effect in the criterion with the partner effect in the judgment *response-set accuracy.*

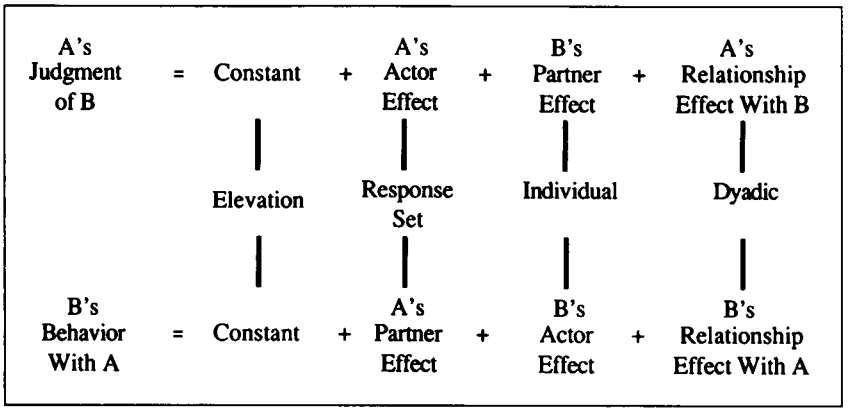

Figure 3. Four nomothetic types of accuracy.

thetic. They measure accuracy for a given trait, as opposed to accuracy for a given judge. The model is also componential. The judgment and criterion are divided into components, and accuracy is measured by the correspondence between sets of components. Finally, it is interpersonal. A person can be both judge and target. The social relations model explicitly recognizes the two-sided nature of social interaction and can measure mutuality of social perceptions.

Design and Measurement Considerations

A number of somewhat technical issues must be considered if one is to develop a new approach to the measurement of accuracy. They are the criterion data structure, choice of criterion measure, research design, and analysis strategy. As will be seen, these choices made by the researcher are closely tied.

Criterion Data Structure

If there is a set of judges who rate targets on a given trait, the judgments result in a two-way matrix of judge by target. The criterion can be in one of three forms. First, it can vary for each judge-by-target combination, and so it, too, is a judge-by-target matrix. Examples of a judge-by-target criterion are the behaviors of the target when interacting with the judge and the target's rating of how much he or she likes the judge. We discussed this type of criterion in the previous section.

In the second form, the criterion for a given target on a specific trait is the same for each judge, and so the criterion simply varies by target. As an example of a target criterion, for each target there is a criterion score from a personality test. This is the criterion data structure used by early accuracy researchers. One common type of target criterion is a set of self-ratings.

In the third form, the criterion is the same value for all judges and targets. For instance, all targets are instructed to tell a lie, and the judges are asked to identify who is lying and who is telling the truth. One can call this a *constant criterion.* Examples of these three types of criteria are illustrated in Table 2.

Criterion Measure

Problematic in any study of the accuracy of social perception is the choice of criterion measure. Because the topic is social judgment, how can one establish what the correct answer is? In the literature, there are five basic types of criterion scores: self-report judgments, third-person judgments, objective measurements, mean judge ratings, and operational criteria.

A self-report criterion is one in which the target is asked to provide the criterion score. The validity of self-report criterion is certainly open to question when the criterion is personality self-reports. For some questions, however, especially those with a judge-by-target data structure, target judgments can be considered perfectly valid. For instance, if judges are asked to guess the extent to which the target likes the judge, the validity of the criterion will be high.

Third-person judgments are expert judgments or those of some knowledgeable informant, such as the spouse. They normally result in a target criterion.

Objective measures are those variables that are measured with little or no human judgment. Examples include percentage of time spent speaking and such physiological variables as heart rate.

Ironically, the judgments themselves are sometimes used as the criterion. For instance, mean judge rating is sometimes taken as the criterion. The reasoning is that if judges are, in general, accurate, they should agree with each other. But consensus, or agreement (what we earlier called *partner variance*), is a necessary but not sufficient condition for accuracy. The use of mean judgment as a criterion for accuracy should not be the method of choice. Some external criterion measure is generally preferable. (Less frequently, the target's judgment of the judge can be used as a criterion. If, for instance, the judgment task is

Table 2
Examples of the Three Criterion Data Structures

Criterion	Example
Judge by target	Impression studies
	Perception of liking
	Self-disclosure
	Frequency of interaction
Target	Personality-judgment studies
Constant	Lie-detection studies
	Pose-emotion studies

to rate the relationship—e.g., "How long have you and the target known each other?"—the target and judge should agree in their ratings of each other. In this case, accuracy is indicated by reciprocity.)

Finally, there is what can be called an *operational criterion.* For example, all targets are instructed to lie (Bond et al., 1985), or subjects are asked to describe someone that they dislike (DePaulo & Rosenthal, 1979). The status of the criterion is established by the task itself. This type of criterion results in a constant criterion data structure.

Design

An important issue concerns the arrangement of judges and targets. The major alternatives follow: (a) Everyone serves as judge and target; (b) one group of people serves as judges, and another group serves as targets; or (c) the judges are unique for each target (or, less commonly, the targets are unique for each judge).

The three designs are diagrammed in Table 3. The first design is called *reciprocal* because for a given judgment each person serves as both a judge and a target. The second design, in which for a given judgment each person is either a judge or a target, is called *classic* because it represents how social and personality psychologists normally study social perception. Although most designs are what are here called *classic,* important studies use the reciprocal design. For example, the data used by Cronbach (1955) in his critique are reciprocal, not classic. The final design is called *nested* because a given target is judged by a unique set of judges.

Each of these three designs has advantages and disadvantages. Because judges and targets are the same people, the reciprocal design ensures that judges and targets are comparable. But because each judge knows that he or she is also a target, he or she may feel evaluation apprehension. One of the most important features of the reciprocal design is that it does capture the interactive, two-sided nature of social perception. For instance, one can measure the degree to which if A sees B high on the trait, B sees A high on the trait. Measures of mutuality and congruence are not possible with the other two designs. Measuring mutuality and congruence can be important in determining whether they mediate accuracy.

The classic design is easy to implement, especially if one needs detailed information about the targets. When the targets are prepared stimuli, as in Buck's (1984) CARAT or Archer and Akert's (1977) measure, this design is implicitly being used. This design does not, however, capture the interactive nature of interpersonal perception.

The nested design is fairly common in the literature (e.g., Woodruffe, 1984). Subjects are asked to list acquaintances, and these people's judgments are solicited. One problem is that judges and targets are confounded because each judge evaluates only a single target. We show in the Data Analysis section that this can lead to serious interpretative problems.

Table 3
Three Designs Used in Accuracy Studies

	Target			
Judge	1	2	3	4
	Reciprocal			
1		X	X	X
2	X		X	X
3	X	X		X
4	X	X	X	
	Target			
Judge	4	5	6	
	Classic			
1	X	X	X	
2	X	X	X	
3	X	X	X	
	Target			
Judge	7	8	9	
	Nested			
1	X			
2	X			
3		X		
4		X		
5			X	
6			X	

Data Analysis

We have discussed three types of designs—reciprocal, classic, and nested—and three types of criterion data structures—judge by target, target, and constant. The analysis of the data depends on the particular combination of the design and criterion data structure used. We consider the analysis of the reciprocal design first. There are two variants of the reciprocal design: round robin and block. In the round-robin design, each judge evaluates each target. This pattern is illustrated in the top of Table 3. In the block design, the people are divided into two groups. The people in one group judge all members of the other group. For instance, DePaulo et al. (1987) created seven groups of six persons. The six persons were numbered 1 through 6. Persons 1, 3, and 5 judged and were judged by Persons 2, 4, and 6.

We first consider the analysis of the reciprocal design with a judge-by-target criterion. Such a criterion was assumed in Table 1 and Figure 3. The social relations model can be used to estimate the four types of accuracy presented in Figure 3 (DePaulo et al., 1987; Kenny, 1981; Miller & Kenny, 1986). Mutuality and congruence effects can also be estimated, and whether they mediate accuracy effects can be determined.

We now consider the analysis by using the social relations model of data in which the criterion does not change for each judge, a target criterion. Given this type of criterion, one can measure the amount of partner variance in the judgments. Partner variance is a measure of consensus and is a necessary condition for individual accuracy. If there is partner variance, the partner effect in the judgments can be correlated with the criterion for each trait to obtain a measure of individual accuracy.

Measures of dyadic accuracy (as well as response-set accuracy), however, are not possible.

When the criterion does not change for each judge or for each target (a constant criterion), response-set, individual, and dyadic accuracy cannot be measured. Only a form of elevation accuracy is possible. The mean judgment can be compared with chance responding. For instance, Bond et al. (1985) asked targets to tell a lie and videotaped these descriptions. Judges then viewed the videotapes and guessed whether the targets were lying or telling the truth. Bond et al. examined the mean of judgments and noted that the judges' likelihood of saying the targets were lying was above chance.

If there are multiple traits, it is possible to measure the correspondence of the judgment to the criterion across traits for each judge–target combination. For instance, Lanzetta and Kleck (1970) measured the number of times a judge correctly guessed whether a target was being shocked or not. If the design is classic or reciprocal, the resulting data structure is a two-way matrix of accuracy scores, judge by target. This two-way judge-by-target matrix of accuracy scores can be analyzed by the social relations model. The actor variance from such an analysis measures the extent to which some people are better judges of the targets than others. The partner variance measures the extent to which some targets are more accurately judged than others. The relationship variance measures the unique ability of a judge to be especially accurate with a particular target. Given a reciprocal design, the social relations model measures the extent to which highly accurate judges are also targets that are judged highly accurately. Additionally measured is the correlation between a judge's ability at judging a given target and that target's being particularly good at judging that judge. The former correlation is the actor–partner correlation, and the latter is the relationship correlation (Sabatelli, Buck, & Kenny, 1986).

A problematic issue is how to measure the correspondence between the judgment and the criterion across traits. If judgment and criterion are dichotomies, it is common to measure percentage correct. Signal-detection measures, however, are generally more appropriate (Lord, 1985; Swets, 1986). If both the judgment and the criterion are continuous in nature, some measure of profile similarity is needed. The correlation coefficient is the usual measure and has the advantage of expressing the amount of accuracy in a scale-free metric, but it is sensitive to the variance of the judgment and the criterion. The regression coefficient has the advantage of being less affected by changes in the variance, but because it has no upper limit, it cannot be claimed that people with higher slopes are more accurate. (Regardless of whether a regression or a correlation coefficient is used, one should cautiously interpret the overall mean because it may indicate just stereotype accuracy.) A profile distance measure, d^2, has numerous problems, as Cronbach (1955) pointed out.

Considerations of the analysis of the classic design are essentially the same as those of the analysis of the reciprocal design. The major limitation of the classic design is that measures of mutuality and congruence are not possible. The model is a bivariate (judgment and criterion), two-way (judge by target) random-effects model. Although such models have been presented in the literature (Abelson, 1960; Cronbach, Gleser, Nanda, & Rajaratnam, 1972; Norman, 1967), they do not appear to have been applied to the accuracy question before.

There has been considerable interest in the analysis of the nested design among clinical psychologists. Ordinarily, clients are nested within therapists. Stiles and Snow (1984) developed component measures of accuracy within a clinical setting. They developed measures of therapist and client accuracy that correspond to our measures individual and dyadic accuracy. Their approach is a multivariate generalization of the univariate model proposed by Howard, Orlinsky, and Perilstein (1976). But, whereas the Howard et al. model is random, the Stiles and Snow model is fixed.[6]

A random-effects model for the nested design can be estimated as follows: First, the intraclass correlation can be used to measure agreement or consensus (Kenny & La Voie, 1985). The intraclass correlation measures the extent to which two judges agree more when rating a common target than when rating a different target. The mean rating of a target can then be correlated with the target's criterion score. If there is a judge-by-target criterion, Kenny and La Voie's (1985) measures can be used to measure individual and dyadic accuracy. (Ironically, individual accuracy is indexed by a group correlation and dyadic accuracy, by the individual-level correlation. The reason for this is that each target has a "group" of judges.)

Because judges are ordinarily not randomly assigned to targets in the nested design, judges of the same target are probably more similar to one another than are judges of different targets. In the parlance of the social relations model, the judges will have similar constants. The nested design is likely to result in inflated estimates of consensus.[7] It can also be shown that individual accuracy is confounded with elevation and that dyadic accuracy is confounded with response-set accuracy. One must then cautiously interpret measures of consensus and accuracy from the nested design.

Given our premise that accuracy research should be nomothetic, interpersonal and componential, we believe that the ideal study should use a reciprocal design with a judge-by-target criterion. We recognize, however, that if accuracy studies were restricted to the reciprocal design, many interesting accuracy issues might be precluded. For instance, in contexts in which judges are naturally nested within targets (i.e., a client–therapist context), a reciprocal design may not be possible.

Example

Anderson (1984) collected the data used for this analysis in his study of perception of social self. Anderson was interested in studying how accurate people are in predicting their "social selves," that is, how accurate people are in knowing how others perceive them. His subjects were members of three fraternities and two sororities. The groups ranged in number from 16 to

[6] Both Howard, Orlinsky, and Perilstein (1976) and Stiles and Snow (1984) treated session as a factor nested within clients. It is likely that the session effects show serial dependency, which complicates the analysis. One strategy might be to treat session as fixed and client and therapist as random. Correlations across sessions would then be estimated within client but tested by using client as the unit of analysis.

[7] Funder (1980, p. 481) also raised this objection to the nested design.

31. Each member of each group was asked to (a) rank order each member of her or his group (including self) on four variables and (b) predict the rank that each member of the group would assign him or her. The first is referred to as the *impression* and the second, as the *prediction.*

The variables on which each subject was asked to rank the others and estimate the partner's ranking of him or her were intelligence, humor, considerateness, and defensiveness. Subjects were given sheets with numbered blanks according to the group size. If there were 16 members in the group, for instance, each was given a sheet containing 16 numbered blanks on which he or she was to fill in the names of the members of the group. For example, for the intelligence variable, the subject was instructed to write the name of the most intelligent person in the group (including the subject) in the first blank and the name of the least intelligent person in the last blank. The subject was then asked to write the name of the second most intelligent person in the second blank and to continue in this manner until all the blanks were filled. The subject was instructed to use the entire range of ranks, assigning only one person to each rank.

In addition, subjects were asked to predict the ranks that each member of the group would assign to him or her. In this case, however, the range was not controlled; the subject could predict a given rank more than once and was not required to use the entire range of ranks. A subject could predict, for example, three people assigning him the rank of fourth. Accuracy was highly stressed in this study; subjects were requested to be as accurate as possible when making their ratings and their predictions. Finally, we converted all scores into normalized ranks.

We must make one note concerning the data. Because of a clerical error, the humor variable was not obtained for one of the sororities. The correlations for this variable then are based on four groups and 96 subjects; the other variables are based on five groups and 121 subjects.

We used this data set to study accuracy in two ways. First, we used the impressions as the criterion and the predictions as the judgments. The question here was, Do people know how others see them? Because the criterion formed a judge-by-target data structure, measures of individual and dyadic accuracy could be computed. Second, we could treat the impressions as the judgments and the self-perceptions as the criterion. The question was, Do peers view the person as self does? Because the criterion was the same for each judge, only individual accuracy could be assessed.

We analyzed the data by using the computer program SOREMO (Kenny, 1987). SOREMO, an acronym for social relations model, partitions variance into actor, partner, and relationship components and provides estimates of correlations between actor and partner effects and between relationship effects for round-robin data structures. All correlations presented in this article are statistically significant at the .05 level of significance.

Results

We first need to consider the variance partitioning. In Table 4, the impressions and predictions are partitioned into actor, partner, and relationship sources. Consider the impression variables. There is no actor variance because subjects rank ordered the other members in the group. The mean normalized rank score must then be 0 for each judge. There is substantial partner variance in the impression ranking. Thus, there is consensus in how these people are viewed by their peers. Whereas an average of 33% partner variance may seem small, the reliability of the impressions for 21 judges is .90. There is also substantial relationship variance.

Table 4
Variance Partitioning for Anderson (1984) Study

Trait	Actor[a]	Partner	Relationship
	Impression		
Intelligent	—	35	65
Humorous	—	48	52
Defensive	—	22	78
Considerate	—	28	72
	Prediction		
Intelligent	43	2	55
Humorous	51	3	46
Defensive	35	3	62
Considerate	31	5	64

Note. Entries are percentages of total variance.
[a] Actor variance of impressions is forced to be zero because of the ranking procedure used.

The variance partitioning of the prediction of impressions is also presented in Table 4. There is substantial actor variance in predictions, averaging 40%. Some people believe that they make a good impression, and others believe that they make a poor impression. The partner variance in predictions is small, averaging 3%. There is a slight tendency for some people to be seen as harsh judges and others to be seen as lenient ones.

As described in a previous section of this article, accuracy is measured by the correspondence between the impression of the actor and the actor's prediction of the partner's ratings. With a social relations analysis of a judge-by-target criterion, one can obtain measures of the four types of accuracy—elevation, response-set, individual, and dyadic accuracy. Elevation accuracy represents the correspondence between the mean level of the impressions and the mean level of the predictions. Response-set accuracy measures the correlation of the actor effect in impressions with the partner effect in predictions. Because there was no actor variance, this component could not be measured. Individual-level accuracy, based on the correlation between the actor component of the predictions and the partner component of the impressions, reflects the overall level of accuracy. Do people know how they are generally perceived across individuals? Dyadic accuracy, based on the correlation between the relationship components of the impressions and predictions, represents the unique level of accuracy. How accurate are people in predicting how a specific partner will rate them?

For elevation accuracy, the mean level of predictions was higher than the mean level of impressions on all four measures. Thus, people felt that they were more favorably perceived than they actually were. As can be seen in Table 5, the individual-accuracy correlations are impressive, averaging in the high .50s. The dyadic-accuracy correlations are much smaller, averaging only .17. The individual-level correlations have no measure-

Table 5
Accuracy Correlations for Anderson (1984) Study

Trait	Accuracy	
	Individual	Dyadic
Intelligent	.594	.112
Humorous	.758	.212
Defensive	.471	.136
Considerate	.475	.225

ment error, however, whereas the dyadic-level correlations do. Even when measurement error is controlled for, the dyadic correlations do not approach the individual-level correlations. If the four variables are treated as indicators of a single trait, individual accuracy is .67, and dyadic accuracy is .46.

One might wonder whether the accuracy results might be mediated by mutuality and congruence effects. For instance, if both Mary and Jane think that each other are intelligent (mutuality) and each then feels that the other thinks that she is intelligent (congruence), then each will accurately know how the other views her (see Figure 2). Because Anderson (1984) used a ranking procedure, mutuality and congruence cannot be measured at the individual level. But such mediation seems relatively implausible at the individual level because mutuality effects at that level appear to be weak in other data (Kenny & Nasby, 1980). The mutuality of impressions and congruence effects at the dyadic level are presented in Table 6. For all four traits, the ordering of the effects is congruence > accuracy > mutuality. For there to be complete mediation of accuracy by congruence and mutuality, mutuality effects must be at least as large as accuracy, which is not the case with this data set.

If one treats the self data as the criterion and the impressions as the judgment, it is possible to examine relations between the self-perception and the peers' perception of the person's standing on the traits. Again for elevation accuracy, people viewed themselves as higher than their peers viewed them. The individual-accuracy correlations, as seen in Table 7, are quite impressive. Correlations between self-perception and peers' perception of intelligence and humor were .401 and .651, respectively, indicating much overlap between how an individual is seen and how he or she sees him- or herself. Peers and self agreed less for the traits defensive and considerate.

Discussion

We addressed two types of accuracy questions. First, do people know how others see them? Second, do peers see people as

Table 6
Dyadic Mutuality and Congruence Correlations for the Anderson (1984) Study

Trait	Mutuality of impressions	Congruence
Intelligent	.098	.130
Humorous	.160	.335
Defensive	.116	.261
Considerate	.166	.352

Table 7
Correlation of Self-Perception With Peer's Perception

Trait	Correlation
Intelligent	.401
Humorous	.651
Defensive	.278
Considerate	.223

people see themselves? Turning to the first question, we found that elevation accuracy was poor because subjects overestimated their standing in the group. Whereas overestimation is the typical result (Myers & Ridl, 1979), underestimation effects have been observed in well-acquainted groups (Israel, 1958). Because a ranking procedure was used, we did not estimate response-set accuracy. Individual accuracy was quite impressive, but dyadic accuracy was not as large. On the basis of this study, we believe that individuals seem to know more accurately how they are generally viewed than they know how specific others view them. This result is consistent with DePaulo et al.'s (1987) review. In well-acquainted groups, individuals seem "to know where they stand" in the group, but they have limited knowledge of how they are differentially viewed by others. Perhaps people believe that others view them basically the same way and fail to appreciate differences in perception. Turning to the second question, we found that peers agreed with self.

One consistent finding is a tendency for larger relations for intelligent and humorous than for defensive and considerate. The former two traits have more partner variance in impression, individual accuracy, dyadic accuracy, and self–peer correlation than the latter two traits. One possible explanation of the difference is that the two traits are more publicly observable (Kenrick & Stringfield, 1980).

Summary

Accuracy is a fundamental topic in person perception. This area has attracted many researchers but has lacked a method with which to measure appropriately and analyze such data. The social relations model provides such a method. We argue that research in accuracy should be nomothetic, interpersonal, and componential.

The social relations model can be used to address the traditional accuracy question, Can people accurately perceive the personality attributes of others? For example, everyone in a group rates one another and is then allowed to interact. The criterion is the measurement of the target's behavior when in the presence of each judge. Studies like this meet the two criteria for accuracy research called for by Funder (1987). Funder argued that accuracy involves the degree to which judges agree with one another and the degree to which these judgments yield valid predictions of behavior. Judge agreement is given by partner variance in the ratings, and behavioral prediction is given by what we call *individual* and *dyadic accuracy*.

We have also suggested a second approach to accuracy, as presented in the example. This approach changes the nature of the accuracy question from the traditional issue to how accu-

rate people are at knowing how others perceive them. Everyone in the group rates each other and also predicts the rating that each member would give her or him. With this approach, the predictions are the judgments, and the ratings are the criteria.

If one is to use "how a person thinks another views him- or herself" as the judgment and "how the person is actually viewed by the other" as the criterion, the accuracy question has been redefined. What is the usual judgment in typical research becomes the criterion. Although the question is not the traditional one in accuracy research, it is still an important one.

We believe that the new wave in accuracy research should concentrate on using naturally occurring stimuli (targets). The use of preselected stimuli has the following drawbacks: First, external validity is seriously compromised. One does not know whether the results are limited to the researchers' stimuli. For example, it is not clear whether the stimuli chosen by Cline and Richards (1960) were selected because subjects were accurate. Second, targets, as well as judge, are an appropriate unit of analysis. By restricting the number of stimuli, appropriate analyses on the targets are precluded. Third, attempts to select targets (Archer & Akert, 1977) encourage individual-difference research. We are not optimistic about such research at this time.

Our approach does have a number of serious drawbacks. These need to be considered before a researcher attempts to use our approach. The major drawbacks that we see are the requirements of multiple partners, possible restrictions in variability, continuous independent and dependent variables, and software availability.

Although the traditional accuracy question concerns the ability of a judge to predict the behavior of multiple targets, in a number of important literatures a judge attempts to predict the behavior of a single target. For instance, a husband predicts the wife's behavior, or a therapist predicts a patient's behavior. The designs of such research can be called simple, dyadic designs, as opposed to the multiple-partner designs presented in Table 3. Within our approach, such designs are variants of the nested design, but all four types of accuracy are confounded. With such data, different analyses might be undertaken (Kenny & Acitelli, 1987).

The second problem with accuracy research from a social relations perspective concerns variance. Because accuracy is measured by correlation, ratings of targets must vary. If all members of the group are fairly similar and, therefore, are rated similarly, accuracy will appear to be low. Thus, subjects must truly vary on the trait for accuracy to be assessed. This problem is present, however, in any correlational approach to accuracy.

We have generally assumed that both the criterion and judgment are measured at the interval level of measurement. If both variables are at the nominal or ordinal level of measurement, the methods that we have described are not appropriate. Wasserman (1987) described an approach to the accuracy question when both variables are at the nominal level of measurement.

The social relations model requires the use of specialized software. Standard statistical packages cannot properly analyze such data. This software has been described by Kenny and La Voie (1984) and by Malloy and Kenny (1986).

The accuracy question is deceptively easy to pose: Do people know what others are like, or do people know how others see them? Cronbach (1955) and others have shown that a naive analysis of this simple question is misleading. One must break the judgment and the criterion into components and measure accuracy in terms of the correspondence between components. In this article, we have proposed a new partitioning of components. We hope that the complicated but necessary approach that we have developed will not have the same effect as the Cronbach (1955) critique. Instead of nipping the rebirth of accuracy research in the bud, we have provided researchers with a framework, albeit a complicated one, with which interpretable and replicable accuracy research can be done.

References

Abelson, R. P. (1960). Scales derived by consideration of variance components in multi-way tables. In H. Gulliksen & S. Messick (Eds.), *Psychological scaling* (pp. 169–186). New York: Wiley.

Anderson, R. D. (1984). *Measuring social-self perception: How accurately can individuals predict how others view them?* Unpublished doctoral dissertation, University of Connecticut, Storrs.

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.

Ausubel, D. P., Schiff, H. M., & Gasser, E. B. (1952). A preliminary study of developmental trends in socioempathy: Accuracy of perception of own status and others' sociometric status. *Child Development, 23,* 111–128.

Bond, C. F., Jr., Kahler, K. N., & Paolicelli, L. M. (1985). The miscommunication of deception: An adaptive perspective. *Journal of Experimental Social Psychology, 21,* 331–345.

Bronfenbrenner, U., Harding, J., & Gallwey, M. (1958). The measurement of skill in social perception. In D. C. McClelland, A. L. Baldwin, U. Bronfenbrenner, & F. L. Strodtbeck (Eds.), *Talent and society* (pp. 29–111). Princeton, NJ: Van Nostrand.

Buck, R. (1984). *The communication of emotion.* New York: Guilford.

Cline, V. B. (1964). Interpersonal perception. In B. Maher (Ed.), *Progress in experimental personality research* (Vol. 1, pp. 221–284). New York: Academic Press.

Cline, V. B., & Richards, J. M., Jr. (1960). Accuracy in interpersonal perception: A general trait? *Journal of Abnormal and Social Psychology, 60,* 1–7.

Cook, M. (1979). *Perceiving others: The psychology of interpersonal perception.* London: Methuen.

Cronbach, L. J. (1955). Processes affecting scores on "understanding of others" and "assumed similarity." *Psychological Bulletin, 52,* 177–193.

Cronbach, L. J. (1958). Proposals leading to analytic treatment of social perception scores. In R. Tagiuri & L. Petrullo (Eds.), *Person perception and interpersonal behavior* (pp. 353–379). Stanford, CA: Stanford University Press.

Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles.* New York: Wiley.

Crow, W. J., & Hammond, K. R. (1957). The generality of accuracy and response set in interpersonal perception. *Journal of Abnormal and Social Psychology, 54,* 384–390.

DePaulo, B. M., Kenny, D. A., Hoover, C. W., Webb, W., & Oliver, P. V. (1987). Accuracy of person perception: Do people know what kind of impressions they convey? *Journal of Personality and Social Psychology, 52,* 303–315.

DePaulo, B. M., & Rosenthal, R. (1979). Telling lies. *Journal of Personality and Social Psychology, 37,* 1713–1722.

Funder, D. C. (1980). On seeing ourselves as others see us: Self–other agreement and discrepancy in personality ratings. *Journal of Personality, 48,* 473–493.

Funder, D. C. (1987). Errors and mistakes: Evaluating the accuracy of social judgment. *Psychological Bulletin, 101,* 75–90.

Gage, N. L., & Cronbach, L. J. (1955). Conceptual and methodological problems in interpersonal perception. *Psychological Review, 62,* 411–422.

Gage, N. L., Leavitt, G. S., & Stone, G. C. (1956). The intermediary key in the analysis of interpersonal perception. *Psychological Bulletin, 53,* 258–266.

Harackiewicz, J., & DePaulo, B. M. (1982). Accuracy of person perception. A component analysis according to Cronbach. *Personality and Social Psychology Bulletin, 8,* 247–256.

Hastie, R., & Rasinski, K. A. (in press). The concept of accuracy in social judgment. In D. Bar-Tal & A. Kruglanski (Eds.), *The social psychology of knowledge.* Cambridge, England: Cambridge University Press.

Higgins, E. T., Herman, C. P., & Zanna, M. P. (Eds.). (1981). *Social cognition: The Ontario Symposium* (Vol. 1). Hillsdale, NJ: Erlbaum.

Howard, K. I., Orlinsky, D. E., & Perilstein, J. (1976). Contribution of therapists to patients' experiences in psychotherapy: A components of variance model for analyzing process data. *Journal of Consulting and Clinical Psychology, 44,* 520–526.

Israel, J. (1958). Self-evaluation in groups. *Acta Sociologica, 3,* 29–47.

Kenny, D. A. (1981). Interpersonal perception: A multivariate round robin analysis. In M. B. Brewer & B. E. Collins (Eds.), *Scientific inquiry in the social sciences* (pp. 288–309). San Francisco: Jossey-Bass.

Kenny, D. A. (1986). Methods for measuring dyads and groups. In W. Crano & M. B. Brewer (Eds.), *Principles and methods of social research* (pp. 288–309). Boston: Allyn-Bacon.

Kenny, D. A. (1987). *SOREMO: A FORTRAN program for round robin data structures.* Unpublished manuscript, University of Connecticut.

Kenny, D. A., & Acitelli, L. K. (1987). *Spouses and other strangers: A round robin analysis of similarity and agreement.* Unpublished manuscript, University of Connecticut.

Kenny, D. A., & La Voie, L. (1984). The social relations model. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 18, pp. 142–182). Orlando, FL: Academic Press.

Kenny, D. A., & La Voie, L. (1985). Separating individual and group effects. *Journal of Personality and Social Psychology, 48,* 339–348.

Kenny, D. A., & Nasby, W. (1980). Splitting the reciprocity correlation. *Journal of Personality and Social Psychology, 38,* 249–256.

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.

Kruglanski, A. W., & Ajzen, I. (1983). Bias and error in human judgment. *European Journal of Social Psychology, 13,* 1–44.

Lanzetta, J. F., & Kleck, R. E. (1970). The encoding and decoding of nonverbal affects in humans. *Journal of Personality and Social Psychology, 16,* 12–19.

Lord, R. G. (1985). Accuracy in behavioral management: An alternative definition of rater's cognitive schema and signal detection theory. *Journal of Applied Psychology, 70,* 66–71.

Malloy, T. E., & Kenny, D. A. (1986). The social relations model: An integrative method for personality. *Journal of Personality, 54,* 101–127.

McArthur, L. Z., & Baron, R. M. (1983). Toward an ecological theory of social perception. *Psychological Review, 90,* 215–238.

McHenry, R. (1971). New methods of assessing accuracy in person perception. *Journal of the Theory of Social Behaviour, 1,* 109–119.

Miller, L. C., & Kenny, D. A. (1986). Reciprocity of self-disclosure at the individual and dyadic levels: A social relations analysis. *Journal of Personality and Social Psychology, 50,* 713–719.

Myers, D. G., & Ridl, J. (1979, August). Self-ratings: Can we all be better than average? *Psychology Today,* pp. 89–98.

Nisbett, R. E., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment.* Englewood Cliffs, NJ: Prentice-Hall.

Norman, W. T. (1967). On estimating psychological relationships: Social desirability and self-report. *Psychological Bulletin, 67,* 273–293.

Richards, J. M., Jr., & Cline, V. B. (1963). Accuracy components in person perception scores and the scoring system as an artifact in investigations of the generality of judging ability. *Psychological Reports, 12,* 363–373.

Rosenthal, R., Hall, J. A., DiMatteo, M. R., Rogers, P. L., & Archer, D. (1979). *Sensitivity in nonverbal communication: The PONS test.* Baltimore, MD: Johns Hopkins University Press.

Sabatelli, R. M., Buck, R., & Kenny, D. A. (1986). A social relations analysis of nonverbal communication accuracy in married couples. *Journal of Personality, 54,* 513–527.

Schneider, D. J., Hastorf, A. H., & Ellsworth, P. C. (1979). *Person perception* (2nd ed.). Reading, MA: Addison-Wesley.

Smith, H. C. (1966). *Sensitivity to people.* New York: McGraw-Hill.

Stiles, W. B., & Snow, J. S. (1984). Counseling session impact as viewed by novice counselors and their clients. *Journal of Counseling Psychology, 31,* 3–12.

Swann, W. B., Jr. (1984). Quest for accuracy in person perception: A matter of pragmatics. *Psychological Review, 91,* 457–477.

Swets, J. A. (1986). Indices of discrimination or diagnostic accuracy: Their ROCs and implied models. *Psychological Bulletin, 99,* 100–117.

Tagiuri, R. (1969). Person perception. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology* (Vol. 3, 2nd ed., pp. 395–449). Reading, MA: Addison-Wesley.

Tagiuri, R., Bruner, J. S., & Blake, R. B. (1958). On the relation between feelings and perception of feelings in small groups. In E. E. Maccoby, T. M. Newcomb, & E. L. Hartley (Eds.), *Readings in social psychology* (3rd ed., pp. 110–116). New York: Holt, Rinehart & Winston.

Warner, R. M., Kenny, D. A., & Stoto, M. (1979). A new round robin analysis of variance for social interaction data. *Journal of Personality and Social Psychology, 37,* 1742–1757.

Wasserman, S. (1987). Conformity of two sociometric relations. *Psychometrika, 52,* 3–18.

Woodruffe, C. (1984). The consistency of presented personality: Additional evidence from aggregation. *Journal of Personality, 52,* 307–317.

Zuckerman, M., Koestner, R., & Alton, A. O. (1984). Learning to detect deception. *Journal of Personality and Social Psychology, 46,* 519–528.

Received May 28, 1986
Revision received March 30, 1987
Accepted March 30, 1987 ■