

Errors and Mistakes: Evaluating the Accuracy of Social Judgment

David C. Funder
University of Illinois at Urbana-Champaign

Laboratory research on "error" in social judgment has largely supplanted research that addresses accuracy issues more directly. Moreover, this research attracts a great deal of attention because of what many take to be its dismal implications for the accuracy of human social reasoning. These implications are illusory, however, because an error is not the same thing as a "mistake." An error is a judgment of an experimental stimulus that departs from a model of the judgment process. If this model is normative, then the error can be said to represent an incorrect judgment. A mistake, by contrast, is an incorrect judgment of a real-world stimulus and therefore more difficult to determine. Although errors can be highly informative about the *process* of judgment in general, they are not necessarily relevant to the *content* or accuracy of particular judgments, because errors in a laboratory may not be mistakes with respect to a broader, more realistic frame of reference and the processes that produce such errors might lead to correct decisions and adaptive outcomes in real life. Several examples are described in this article. Accuracy issues cannot be addressed by research that concentrates on demonstrating error in relation to artificial stimuli, but only by research that uses external, realistic criteria for accuracy. These criteria might include the degree to which judgments agree with each other and yield valid predictions of behavior.

The accuracy of human social judgment is a topic of obvious interest and importance. It is only natural to wonder to what degree the judgments we make of the personalities of ourselves and others might be right or wrong, and to desire to improve our accuracy. Nonpsychologists are often surprised and disappointed, therefore, when they begin to take psychology courses and discover that the field has largely foresworn interest in the accuracy issue. The discipline's early, direct interest in accuracy (e.g., Estes, 1938; Taft, 1955; Vernon, 1933) was all but completely stifled some years ago by the publication of a series of methodological critiques by Cronbach (1955), Hastorf and Bender (1952), and others. As a result, according to one authoritative textbook:

The accuracy issue has all but faded from view in recent years, at least for personality judgments. There is not much present interest in questions about whether people are accurate. . . . There is, in short, almost no concern with normative questions of accuracy. On the other hand, in recent years there has been a renewed interest in how, why, and in what circumstances people are inaccurate. (Schneider, Hastorf, & Ellsworth, 1979, p. 224; see also Cook, 1984)¹

Specifically, the psychology of social judgment has been dominated in recent years by a flood of research on the subject of "error." Studies of error appear in the literature at a prodigious rate, are disproportionately likely to be cited (Christensen-Sza-

lanski & Beach, 1984), and fill whole books (Kahneman, Slovic, & Tversky, 1982; Nisbett & Ross, 1980). Psychology instructors have found that the various experimental demonstrations of error, like other counterintuitive phenomena, provide a sure way to spice up an undergraduate lecture in introductory or social psychology.

The study of error is a "hot topic," therefore, and justifiably so. Errors are an indispensable tool for studying the processes of human judgment. They can provide valuable insights into how the cognitive system transforms, augments, and distorts an initial stimulus "input" on the way to a final judgment "output."

However, there is another, perhaps more important reason why the new research on error is so fashionable. Crandall (1984) writes

Mistakes are fun! Errors in judgment make humorous anecdotes, but good performance does not. It is fun to lean back in our chairs and chuckle about our goofs. (p. 1499)

More seriously, Evans (1984) writes

Although it is nice to know that people are reasoning well or making good decisions in some contexts, it is much more important to know when they are not. . . . Surely, the imperative message for us to impart to decision makers is that of their proneness to error. (pp. 1500-1501)

These quotations exemplify how, in a subtle but unmistakable way, error research has moved well beyond study of the judgment process. Studies of error are equated with studies of whether people "reason well" or "make good decisions." Errors are usually treated as "shortcomings" of judgment, as the title

This research was supported in part by grants from the Clark and Milton Funds of Harvard University and by National Institute of Mental Health Grants R01-MH40808 and R01-MH42427.

I am grateful for the comments of many individuals, including Paul Andreassen, Thomas Carrell, Jean MacMillan, Dan Ozer, Tamra Pearson, Stanton Peele, William Swann, four reviewers, and the Editor.

Correspondence concerning this article should be addressed to David C. Funder, Department of Psychology, University of Illinois, 603 East Daniel Street, Champaign, Illinois 61820.

¹ From *Person Perception* (p. 224) by D. J. Schneider, A. H. Hastorf, and P. C. Ellsworth, 1979, New York: Random House, Inc. Copyright 1979 by Random House, Inc. Reprinted by permission.

of more than one major work on the subject clearly reflects, and the existence of so many errors is taken to indicate that many if not most of the judgments people make about each other are wrong. In this way, by another name, and through the back door, the accuracy issue has reentered social psychology.

This reentry has occurred even though some investigators have occasionally acknowledged that error in the laboratory and inaccuracy in a larger sense may not be the same thing. Some years ago, D. T. Campbell (1959) wrote, "[many] errors . . . will be found to be part-and-parcel of psychological processes of general adaptive usefulness" (p. 340), and Tversky and Kahneman (1983) stated more recently that "the focus on bias and illusion . . . neither assumes nor entails that people are perceptually or cognitively inept" (p. 313).

Such disclaimers have had little impact over the years. Psychology's widespread fascination with error research continues to stem primarily from its apparently dramatic implications for accuracy in daily life (e.g., Crandall, 1984; Evans, 1984), *not* from its value for understanding the mechanisms of judgment. This is unfortunate because, according to the thesis of the present article, research on error is almost completely irrelevant to the accuracy of social judgment, and to the extent that the popularity of error research is due to its apparent implications for accuracy, the emphasis on it has been misplaced. The reason is that an error is not the same thing as a mistake.

An "error" is a judgment of a laboratory stimulus that deviates from a model of how that judgment should be made. When this model is normative and rational, the error represents an incorrect judgment.² A "mistake," by contrast, is an incorrect judgment in the real world, such as a misjudgment of a real person, and so must be determined by different criteria. Detection of an error implies the existence of a mistake *only* when the process that produces the error also produces incorrect judgments in real life. Unfortunately, this cannot be determined by merely demonstrating the error itself, because the same judgment that is wrong in relation to a laboratory stimulus, taken literally, may be right in terms of a wider, more broadly defined social context, and reflect processes that lead to accurate judgments under ordinary circumstances. Most areas of psychology other than the study of social judgment, for example, the study of visual perception, generally assume that errors in the laboratory are the result of mechanisms that produce correct and adaptive judgments in real life. A different sort of research is needed to assess accuracy, therefore. Such research must let subjects judge real people in authentic social contexts, and use realistic, external criteria for determining when these judgments are right or wrong.

The purposes of this article are to summarize the current state of research on accuracy and error in social judgment, to demonstrate how the two kinds of research are not equivalent, and to propose an agenda for research relevant to accuracy issues. The presentation is in several sections. The first section contains a brief history of how research on accuracy came to be supplanted by research on error. The second section covers the difference between errors and mistakes, and illuminates the distinction with reference to research on visual perception. The third section demonstrates the consequences of failing to distinguish between errors and mistakes in the context of three specific examples from the social psychological literature, then outlines some general implications of the distinction for interpret-

ing research on error in social judgment. The fourth section describes in general terms the sort of content-oriented research that *could* address accuracy issues, and the fifth section summarizes some specific research findings. The sixth and final section outlines several possibilities for future research on the accuracy of interpersonal judgment.

Evolution of Research on Accuracy and Error

The attempt to study human error is a remarkable undertaking because it seems to require a criterion, a presumption that the actual, correct state of affairs can be known with certainty. In the study of more purely quantitative judgments this seems to present relatively little difficulty. The odds of having a certain number of coin flips come up "heads" or the probability of a taxicab being a certain color given prior base rates can be calculated with precision using Bayes's theorem (Tversky & Kahneman, 1978). If a judgment problem is framed in such terms, therefore, the correct answer can be exactly determined (although even then not completely without controversy; see Cohen, 1981; Shafer, 1985). Subjects' degree of error can be neatly assessed by the degree to which their estimates diverge from this precise criterion. The availability of a criterion probably helps explain the large volume of research that has been conducted on explicit probability judgments, often in literal terms of odds and percentages.

The translation of this research into prescriptions on how we should make quantitative judgments in daily life has begun to stir controversy, however. As Ebbesen and Konecni (1980) point out, people may not make judgments in the real world the same way they do in the lab. Moreover, explicit, numerical probability judgments are common only within certain specialized occupational groups, such as gamblers and weather forecasters. Even more fundamental is the criticism that although the mathematical rules of probabilistic inference are clear given certain assumptions (e.g., about the nature of outcome distributions and the stability of the underlying process), in real life one can never be sure when these assumptions are valid (Einhorn & Hogarth, 1985; Lanning, 1985). Hogarth (1981) further points out that "continuous," real-world environments give decision makers repeated feedback that sometimes allows correct decisions to be approached incrementally, through repeated trial and error, rather than computed at the onset in a mathematically optimal, Bayesian fashion (cf. Gettys, Kelly, & Peterson, 1973; Lindblom, 1959; Winkler & Murphy, 1973).

When the study is of social judgment, the topic of this article, the criterion problem becomes even stickier. The complexity of most social situations makes any degree of certainty and precision in the establishment of "truth" difficult to come by, and the necessary assumptions are difficult to formulate, much less confirm. If a subject claims that someone is "friendly" or "com-

² Strictly speaking, any departure of the subject's judgment from the model being tested is an "error" even if the model postulates a gravely flawed or irrational judgment process. In such situations the nonevaluative, technical meaning of the term *error* becomes clear. However, such cases are rare. Almost always in the contemporary study of social judgment, the model is of rational judgment, so an error is not only a deviation from it, but also can be considered incorrect (in terms of the immediate stimulus).

petent" on the basis of his or her acquaintance with that person, for example, on what grounds can we assess whether the subject is right or wrong?

From the 1930s into the early 1950s, psychologists did attempt to tackle difficult questions of this sort. Research was conducted that directly assessed the accuracy of subjects' judgments by comparing them to external, more or less realistic criteria. For example, Vernon (1933) examined how well subjects were able to predict the performance of themselves, friends, and strangers on various tests of intelligence, personality, and artistic tendency. Estes (1938) had subjects judge stimulus persons viewed on film, and assessed the accuracy of their judgments by comparing them with judgments made by trained clinicians. Yet the most common sort of criterion, by far, was that used by Dymond (1949, 1950), who assessed the accuracy of judgments made by group members about each other in terms of how well they agreed. By 1955, Taft was able to review a large number of studies of accuracy in interpersonal judgment, mostly of this last sort (see also Schneider et al., 1979).

This lively research tradition came to an abrupt and nearly complete halt after the publication of Cronbach's (1955) methodological critique (see also Gage & Cronbach, 1955). Cronbach demonstrated how the criteria used in these studies, typically profile-similarity scores, were contaminated to an unknown degree by influences such as "stereotypic accuracy," "elevation," and "differential elevation" (for a discussion of these see Funder, 1980, pp. 479-482). The article's criticisms were expressed in a difficult mathematical style and appeared insurmountable to many psychologists.

When examined closely, however, Cronbach's analysis did not need to be taken as implying that research on accuracy is impossible, nor even especially difficult (cf. Funder, 1980). Rather, it addressed some complexities of data interpretation that most researchers had neglected up to that time. Most centrally, it highlighted the necessity of keeping separate the components of judgment due to various nonsubstantive response styles from that aspect of judgment truly relevant to accuracy. A few investigators took heed of his suggestions and pressed on. Most notably, industrial psychologists, whose work is considered later in this article, maintained an interest in accuracy issues. Other studies were contributed over the years by Bernstein and Davis (1982), DePaulo (1978), Harackiewicz and DePaulo (1982), Jackson (1982), and Pryor, Gibbons, Wicklund, Fazio, and Hood (1977).

In general, however, research on accuracy and the *content* of judgments fell off dramatically, and attention began to turn instead to studies of the *process* of judgment that could seemingly bypass or finesse the difficulties entailed by external criteria.³ This shift in emphasis was welcomed, even seen as overdue, by psychologists who regarded the process of judgment as more interesting in the first place. For example, Jones (1985) writes

Despite the obvious importance to social psychology of knowledge about the person perception process, the development of such knowledge was delayed [until the mid-1950s] by a preoccupation with the accuracy of judgments about personality. (p. 87)⁴

Psychologists soon found that research conducted with relative ease in the laboratory, using wholly artificial stimuli, could provide valuable insights into the judgment process without raising any need to be concerned about accuracy issues. Asch

(1946), for example, was able to present a theory of the way people combine social information based solely on data about how subjects evaluate lists of personality trait terms. As Jones (1985) points out, "Asch solved the accuracy problem by bypassing it" (p. 87). Other psychologists interested in the process of social judgment, inspired by Heider (1958), concentrated on building idealized, normative models of the judgment process, such as the "correspondent inference theory" of Jones and Davis (1965), the "attributional cube" of Kelley (1967), and the more general work by Peterson and Beach (1967). These models emphasized the logical structure of judgment, and were based on the working assumption that people are perfectly "rational" processors of information. More recently, researchers have imported methods of cognitive psychology into the study of person perception, creating the burgeoning new field of "social cognition" (e.g., S. T. Fiske & Taylor, 1984; Hastie et al., 1980). This research continues the tradition of using contrived stimuli in artificial, experimental settings to test various models of the person perception process. The use of artificial settings is perfectly appropriate, of course, to theoretical model testing of this sort (cf. Mook, 1983).

Yet such research does not and was never originally intended to address the external validity or *accuracy* of personality judgment. According to M. Ross and Fletcher (1985), "Most attribution theorists are not concerned with the validity of the attributions that the layperson produces. Instead, theorists focus on the cognitive processes involved in forming attributions" (pp. 73-74).⁵

Nevertheless, an interesting thing happened when psychologists began conducting empirical, process-oriented research based on idealized, normative models such as those by Jones and Davis (1965) and Kelley (1967). At first, a few studies supported these models as adequate descriptions of the judgment process when applied to experimental stimuli (e.g., Kelley, 1973; McArthur, 1972, 1976). Gradually, however, it became apparent that actual human judgment often deviates from their prescriptions. Nisbett and Ross (1980), for example, were led to conclude that "we have become increasingly more impressed with the evidence of people's departures from normative standards of inference [attribution models] and less impressed with the evidence of their adherence to them" (p. 6). Simultaneously, these models came to be viewed less as *theories* of how people *do* make judgments and more as *standards* prescribing how people *should* make judgments. As a result, any departure from these models or, more generally, any transformation, biased recall, or other distortion of experimental stimuli began to be taken as an "error," or even less ambiguously, a "shortcoming" or "fallacy"

³ Cook (1984) makes the interesting claim that direct research on accuracy "has become unfashionable partly because apparently simple techniques proved to be methodologically complex, but also because talking of 'accuracy' of perception implies a reality to be perceived, and the current resurgence of phenomenological approaches to social psychology tends to deny any such reality" (p. ix).

⁴ From *The Handbook of Social Psychology* (3rd ed., p. 87) by G. Lindzey and E. Aronson (Eds.), 1985, New York: Random House, Inc. Copyright 1985 by Random House, Inc. Reprinted by permission.

⁵ From *The Handbook of Social Psychology* (3rd ed., pp. 73-74) by G. Lindzey and E. Aronson (Eds.), 1985, New York: Random House, Inc. Copyright 1985 by Random House, Inc. Reprinted by permission.

(cf. E. E. Jones, 1985, p. 92; Nisbett & Ross, 1980; L. Ross, 1977; Tversky & Kahneman, 1983). The final step after that, of inferring these errors to have dismal implications for accuracy in real life, seemed so natural and obvious that apparently it was almost invisible.

The clearest as well as best known example concerns the "fundamental attribution error," a term that seems to have become firmly entrenched in the literature of social psychology. This error is the putative tendency for people to overestimate the influence of attitudes and personality on behavior, and to underestimate the power of the situation.⁶ It is typically demonstrated by experiments that show how subjects will draw inferences about the personalities of rating targets on the basis of insufficient information. Discussions of this error have been relentlessly pessimistic; none of its presentations (e.g., Nisbett & Ross, 1980; L. Ross, 1977) leave any doubt that it may as well have been called the "fundamental attribution mistake." The only slim ray of hope offered is that perhaps people could be trained not to make it (Nisbett & Ross, 1980, chap. 12). This error, like many others, is presumed to reflect a flaw in judgment that people would be better off without.⁷

The history of recent research on social judgment concludes with an ironic twist. A line of investigation initiated with the specific aim of *bypassing* accuracy issues became psychology's principal tool for making normative prescriptions. Subjects were deemed wrong not because their judgments were incorrect in terms of external social reality, but because within the "partial reality" of the laboratory their judgments distorted *experimental* stimuli by failing to follow various rules of logical inference, such as those supplied by normative models of the attribution process. Such distortions lead to incorrect judgments in the laboratory and deserve to be called errors. Yet before regarding these errors as implying inaccuracy in real life, we should consider carefully what they might mean in a wider, more realistic social context. This consideration brings us back to the important question that laid neglected while research on accuracy was gradually replaced by research on error: When is an error a mistake?

Two Meanings of "Error"

Like so many other common words (e.g., "reliability," "significance") that psychological jargon has borrowed from ordinary English, the word *error* appears deceptively simple, because its psychological usage is different. In psychology, proper use of the term is technical, not evaluative. An error is any deviation of an individual observation from a standard, such as a sample or population mean, or the output from a predictive model (e.g., a regression). Psychologists attach no negative connotations whatsoever, therefore, to phrases that refer to this sort of deviation, such as "standard error of estimate," "error variance," and "error term."

Still, these phrases can confuse students, because to the non-psychologist an error is simply a bad thing. The distinction between the two usages is nicely demonstrated in *Funk & Wagnall's Standard College Dictionary* (1963, p. 450), in which the first definition of error is "Something done, said or believed incorrectly; a *mistake*" [emphasis added]. The technical sense is in the fifth definition: "The difference between the observed value of a magnitude and the true or mean value." (Of course,

in this context "true" also has a technical meaning.) Errors demonstrated in the laboratory have this second meaning; they represent departures from the experimenter's standard for a "true" response that directly reflects the stimulus. Yet even some psychologists seem to be less aware than they should be of the subtle difference between this sort of error and a mistake.

The difference is subtle because, rather than being discretely different, errors and mistakes anchor opposite ends of a conceptual continuum. At the error end are misjudgments of narrowly defined, artificial stimuli, which represent departures from a normative model. These misjudgments are of *no intrinsic* interest, but can be highly informative about whether the model accurately depicts the process of judgment. Near the other end, labeled "mistake," are misjudgments of more poorly defined, real-life stimuli. These misjudgments are important because of their potentially dangerous consequences. Although it is relatively easy to detect an error, because the nature of the stimulus is precisely known and the normative judgment of it can be modeled with some certainty, it is much more difficult to determine that a judgment, perhaps even the same judgment, is also a mistake, because the criteria must be much broader, and located in the real world.⁸ In theory, errors and mistakes shade into each other, but in practice, as several examples later in this article show, errors and mistakes are often clearly distinguishable.⁹

⁶ The actual status of this error as "fundamental" has not gone unquestioned. See Funder (1982b), Harvey, Town, and Yarkin (1981), and Monson and Snyder (1977).

⁷ A careful reader of the error literature can find occasional disclaimers. For example, Nisbett and Ross (1980) begin their book by acknowledging that "errors . . . often seem traceable to violations of the same inferential rules that underlie people's most impressive successes" (p. xi). Nonetheless, this article is premised on the observations that (a) the very same investigators who write such disclaimers also, elsewhere and frequently, equate "violations of . . . inferential rules" with flaws of judgment (e.g., discussions of the fundamental attribution error); and (b) much of the widespread popularity of error research is a direct consequence of its apparent, but largely illusory, implications for accuracy.

⁸ An analogy can be drawn with American jurisprudence at the appellate level. Appeal of a trial verdict depends not on whether the verdict was right or wrong, but rather on whether correct procedures were used in reaching it. Accordingly, the criteria for granting the appeal are narrow: Was the verdict based on "extraneous" biases or information from outside the courtroom rather than solely on the evidence presented in court? Did the judge cite incorrect law when instructing the jury? The broader criterion of whether the actual content of the verdict ("guilty" or "not guilty") was correct, its external validity, is not at issue. Notice how the narrow criteria that are used are perfectly appropriate to deciding whether or not the correct legal *process* (which can be considered a normative model) was followed, but are in principle irrelevant to the accuracy of the verdict itself.

⁹ A distinction similar to the one between "errors" and "mistakes" in the present article is drawn by Kruglanski and Ajzen (1983) between "bias" and "error": "bias need not result in error, if by the latter term is meant a departure from some accepted criterion of [external] validity" (p. 18). A reviewer of this article suggested that the term *bias* in fact should *replace* the term *error* in all future research, so that the ambiguity of error could give way to a clear distinction between biases and mistakes. Such change in usage would be helpful, but because the present article is devoted to analysis of a past literature that does widely use error, it is used here despite its ambiguity. Also, Kruglanski and Ajzen use bias as a necessary, integral part of any complete process

Errors in the Study of Visual Perception

One domain in which the distinction between errors and mistakes usually is kept clear is the study of visual perception. Illusions and misperceptions are seldom discussed as if they revealed "shortcomings" in the visual system.

A classic example, often seen in introductory psychology textbooks, is a drawing of two lines that converge upward, as if toward a horizon (Figure 1). A line segment is drawn near the bottom, and a second segment of equal length is drawn near the top. The second segment appears, incorrectly, to be longer than the first. (This is the Ponzo or "railroad lines" illusion; Gregory, 1971.) Although the subject's perception is wrong, notice that in nearly every situation *outside* the laboratory it would be correct. Objects near the horizon that leave equivalent retinal images to nearby objects generally *are* larger. Another famous example is the Müller-Lyer illusion, which is that reversed arrowheads seem to lengthen their connecting shaft, whereas normal arrowheads seem to shrink the shaft (Figure 2). As Gregory (1971) points out, the only place in real life in which shapes like the Müller-Lyer figures appear is on the inside or outside edges of a box (or room), contexts in which a near edge *is* smaller than a far edge that leaves the same retinal image.

According to Gregory (1971), an early interpretation of the illusion in Figure 2 was that it reflects a general tendency to overestimate acute angles and underestimate obtuse ones, although no reason for such a strange tendency was ever provided. Yet in 1896 a psychologist named A. Thiery proposed that the illusion is related to the way people correctly use perspective in real life, and this interpretation eventually became predominant. Unlike their colleagues in social psychology, modern psychologists who study vision no longer treat illusions as flaws of judgment (Over, 1968). Rather, they assume that these illusions reveal "information-processing mechanisms that under normal circumstances make the visual world easier to comprehend" (Gregory, 1971, p. 167; cf. Helmholtz, 1903; Dember & Warm, 1979).¹⁰

Production of Errors and Illusions

Gregory (1971) points out one reason why the usually effective visual system can be fooled by experimental illusions:

Pictures [such as Figures 1 and 2] are highly artificial, and present special problems to the perceiving brain. In a sense all pictures are impossible because they have a dual reality. They are seen both as patterns of lines lying on a flat background and as objects depicted in a quite different three-dimensional space. (p. 171)

This dual reality is what keeps visual errors from being visual mistakes. The pairs of stimulus lines in Figures 1 and 2 really are the same length and any judgment otherwise is wrong, so long as the criterion for accuracy is taken very literally as constituting the length of the ink lines on the page. Yet if the figures are considered not as two-dimensional but as representations of three-dimensional stimuli (e.g., as if they were photographs or retinal images), the situation becomes very different—the

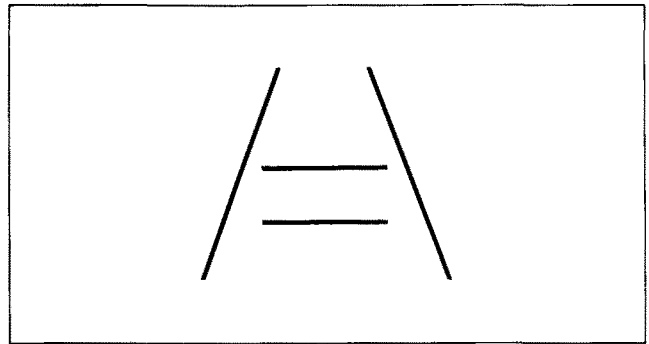


Figure 1. The Ponzo or "railroad lines" illusion: The upper segment appears longer, apparently because interpreted in three dimensions, the more usual perceptual situation, it is nearer the horizon and is in fact longer.

two lines could in fact be of different lengths. If the figures really represent corners or outside edges, or railroad tracks against a horizon, then the two lines *definitely* are of different length, and the perceiver is not wrong. Perceptual psychologists recognize that the three-dimensional case more realistically approximates the usual situation for visual perceivers. Therefore, they waste no time berating their subjects for the error and instead concentrate on the way it reveals how cues are ordinarily combined into correct three-dimensional judgments of distance and size.

Lindsay and Norman (1972) describe another aspect of the difference between experimental illusions and visual reality:

Normally, all contextual information fits together. As objects move away, their image size changes by just the right amount. The relative sizes and distances are what they should be. Neither the artist nor the psychologist, however, is constrained to studying real-life situations. The surrealist delights in discovering and intentionally violating the rules for constructing logical perceptions. (pp. 29–30; cf. Hammond, 1966)

In other words, experimental stimuli in perceptual research can be constructed so as to be not only artificial, but impossible—not in terms of the laboratory, of course (nothing is impossible in a laboratory), but in terms of their extrapolation to wider visual reality. A subject will typically be "fooled" by such stimuli, but there is even less reason than in the three-dimensional case to consider his or her error a mistake. A subject's response to "impossible" stimuli cannot reasonably be taken to imply perceptual incapacity in real life. If such incapacities exist, they will have to be demonstrated in other ways.

Visual psychologists understand these two principles well, but their relevance to social psychology is less often appreciated. First, in the same way that every picture has a dual reality, a social psychological experiment exists to the subject both as a two-dimensional, contrived array of stimuli *and* as a three-

model, rather than as a departure from such a model that yields an incorrect judgment of the proximal stimulus, so their term is not precisely equivalent to error as used in the present article.

¹⁰ This difference may arise because the visual system is generally well regarded; visual psychologists are willing to grant that it works well in ordinary circumstances, and so they take errors as clues about how it works rather than as evidence that it works badly. Some social psychologists, by contrast, seem more pessimistic. If one's overall evaluation of human social perception is that it is badly flawed, then it is more natural (if no more justified) to regard errors as manifestations of these flaws.

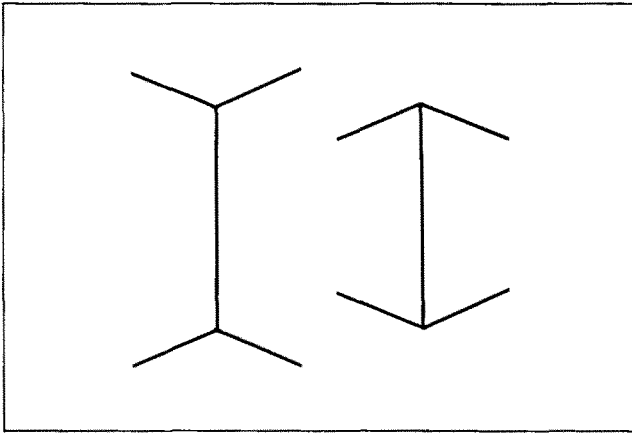


Figure 2. The Müller-Lyer illusion: The left-hand segment appears longer than the right-hand segment, because the arrowheads imply it is a far, inside edge, whereas the arrowheads on the right imply that segment is a nearer, outside edge. (In three-dimensional, real-life settings the processes underlying this illusion generally lead to correct perception.)

dimensional, real situation involving interaction among the subject, the experimenter, and even the institution in which the research takes place (cf. Orne, 1962). Thus, a subject's judgment might be an error in terms of the stimuli with which he or she is explicitly presented, but still correct (or at least defensible) in the wider context of the social situation in which the judgment is being made. Second, depending on one's interests (and, perhaps, one's surrealistic inclinations), a researcher can arbitrarily create patterns of social information or even entire social situations that are strange, unusual, or even impossible. A subject's response to such a situation is more likely to represent a (perhaps desperate) "best guess" under bizarre circumstances than a reflection of his or her basic judgmental capacity. The point is that errors can be defined very literally in terms of the immediate concrete stimulus, but before calling that judgment a mistake one needs to take into account the subject's wider social situation and more usual social experience. The next section begins with a detailed consideration of three representative examples.

Social Errors in a Wider Context

The implications of the principles underlying the production of illusions, just outlined, for research on social judgment can be illustrated by closely considering three well-known, more or less representative studies. Each, in a different way, shows how an error might be confused with a mistake. Yet these studies are not exceptionally flawed, and most of the comments made about them also apply to many others of the same general type. The purpose of discussing these particular studies is to make a general point in the context of some specific examples, and to survey the recent literature on social judgment in a representative rather than exhaustive manner.

Overattribution

In a classic study of the overattribution effect, Jones and Harris (1967) showed subjects essays favoring Castro that were pur-

portedly written by people who had no choice in what to write. This study revealed a tendency for subjects to see such essays as reflecting their authors' true positions anyway. Subjects assumed that pro-Castro essays were written by pro-Castro authors, despite the situational constraint, and thereby committed an attribution error. A large literature has subsequently demonstrated this effect in a variety of ways (for a review see Jones, 1979), and it is usually interpreted as a judgmental flaw. For example, Nisbett and Ross (1980) describe the effect as demonstrating "peoples' [general] overwillingness to ascribe behavior to enduring dispositions" (p. 131).

In their own terms overattribution experiments are perfectly straightforward and the error that subjects make is clear. However, considered in a larger frame, as a social interaction between subject and experimenter, many of these experiments have a curious aspect. The experimenter provides the subject with information that is normatively irrelevant to a judgment, then asks him or her to make that very judgment. In Jones and Harris's (1967) and many subsequent studies, subjects are given an essay and then asked to judge the writer's true attitude. The only clues available to the subjects' judgment are (a) their experience with attitudinal baserates and (b) the essay's content. Because subjects were told that the writer was instructed what to write, experimenters presume that the most rational course would be for subjects to judge the writer's true attitude solely from the baserates. They are deemed in error precisely to the degree that they base their judgments on the essay's content instead. However, subjects have a third piece of information, one that is not "officially" part of the experimental stimulus: (c) the experimenter gave them the essay, for their use (cf. Miller & Rorer, 1982).

As subjects implicitly know, and psycholinguists and sociologists have documented, normal (i.e., extraexperimental) social discourse proceeds according to various rules or "contracts." One of the most important of these rules is that the speaker tell the listener what he or she thinks the listener needs to know. The speaker should not give information that he or she thinks the listener already knows or that he or she believes has no relevance to the situation at hand. Moreover, he or she should not ask questions that the listener has no basis for answering. (For a discussion of these rules, see Clark & Clark's [1980] "given-new contract"; or Goffman, 1974; Grice, 1975; Swann, Giuliano, & Wegner, 1982). As social communicators, experimenters in studies of overattribution routinely violate this contract. Subjects are given information that in one way of thinking, is irrelevant to the judgment they must make. Yet they are given the information in precisely the way that in ordinary discourse, relevant information is provided. Furthermore, they are asked a question ("What is the author's true attitude?") that they cannot answer *unless* they assume the essay to be diagnostic. All of this happens to them while they are subjects in an experiment, under the close scrutiny of psychologists. In terms of the experimental stimuli taken very literally, the subject's response is wrong, an error, just like the perception of someone who looks at Figure 1. However, just as one's perception of Figure 1 is not an error if it is taken as three-dimensional, a subject's response to the total social situation of an overattribution experiment seems to make perfect sense. In this wider context, it seems only natural for subjects to make the "error" of basing their judg-

ments on the information they have been given (cf. Adler, 1984), and inappropriate to construe this error as a mistake.

Moreover, although the overattribution effect is usually cited as deriving from the fundamental attribution error (e.g., Nisbett & Ross, 1980), Quattrone (1982) has shown that a similar experiment will lead subjects to "overattribute" to the *situation* if the experimenter merely asks them to use the essay to estimate situational forces that might be present, instead of the writer's "true" attitude. Thus, the direction of overattribution, that is, to the person or situation, seems to depend more on the specific question the experimenter asks than on any more "fundamental" tendency.

Processing Inconsistent Information

In another study, Berman, Read, and Kenny (1983) gave subjects two word descriptions of hypothetical stimulus persons that were either "consistent with expectations" (p. 1215; e.g., friendly and polite) or not (e.g., shy and sociable). The phenomenon of interest was the subjects' subsequent "erroneous" tendency to rate the stimulus person higher on both traits when the pair was consistent, and lower when the pair was inconsistent. Berman et al. interpreted this tendency as a general flaw in judgment, writing that "the findings of the present experiments should be of particular interest to psychologists who use observers or raters in their research . . . [because they show that] ratings can provide a biased picture of social events" (p. 1223).

Yet in light of the dual reality of social psychological experiments, this situation is even more peculiar than those in which "overattribution" is demonstrated. Both one-word descriptions were presented as ratings of real individuals made by "trained psychologists" (p. 1214). In the critical conditions, the two words either directly implied each other (the "consistent" condition) or, almost oxymoronically, seemed to contradict each other (the "inconsistent" condition). For example, whereas the fictional individual "Keith" was described as "friendly" and "polite," "Edward" was said to be both "shy" and "sociable." It should be emphasized that subjects were told, falsely, that these were real individuals.

Consider some definitions from *Funk & Wagnall's Standard College Dictionary* (1963). *Friendly*: "well-disposed, not antagonistic . . . helpful, favorable" (p. 535). *Polite*: "exhibiting . . . a considerate regard for others, courteous . . . a manner of social intercourse that is designed to please" (p. 1045). Given these definitions, it is difficult to imagine someone who is friendly without also being polite, and vice versa. Consider also *shy*: "bashful, reserved" (p. 1246); and *sociable*: "inclined to seek company" (p. 1272). Given these definitions, it is a complex (though not impossible) task to imagine someone who is both shy and sociable.

Recently, Asch and Zukier (1984) investigated how subjects integrate this sort of information about a person when it is congruent as opposed to incongruent. They found that when a stimulus person is described by a pair of congruent traits, each serves to strengthen the other; to be friendly and polite is to be both very friendly and very polite. Yet when the two traits are incongruent, subjects use a variety of sophisticated strategies to integrate them, all of which involve changing or moderating the meaning of at least one of the terms. For example, the meaning of "shy" might be altered to refer only to encounters with the

opposite sex, whereas "sociable" could be taken to refer to encounters with same-sex buddies. To be both shy and sociable, therefore, is neither to be very shy (or shy in all situations), nor very sociable.

Berman et al.'s (1983) study was quite similar to Asch and Zukier's (1984) except that the former investigators' subjects were more restricted in their responses: They could only respond by rating each trait, from memory, as either "high" or "low." Given such limited response options, it seems perfectly reasonable and appropriate for subjects to rate the inconsistent traits, which Asch and Zukier showed lead to highly qualified impressions, as low relative to the consistent traits, shown by Asch and Zukier to produce unqualified impressions held with clarity and confidence. Notice that Berman et al.'s subjects did make a wrong judgment or error in the narrow, technical sense: Their outputs, or judgments, were not a direct function of the experiment's inputs; they did not literally parrot the contents of the words on the stimulus slides. Instead, the subjects appeared to apply their knowledge of the meaning of the terms to their judgments (cf. J. Block, 1977). Thus in that larger semantic context they may have been doing all they could, given the limits on their responses, to avoid being absurd.

Social Roles

One of the most interesting studies of attribution error is by L. Ross, Amabile, and Steinmetz (1977). Compared to most research in this area, the situation it created for subjects was unusually realistic. Pairs of subjects were randomly assigned to the roles of "questioner" and "contestant" and, in a quiz show-like format, the latter tried to answer questions made up by the former. Afterward, the contestant consistently rated himself or herself as inferior to the questioner in general knowledge. In a second study, uninvolved observers watched a realistic reenactment of the same situation, and manifested the same strong tendency to attribute more general knowledge to the arbitrarily assigned occupant of the questioner role. L. Ross et al.'s discussion of this phenomenon was unrelievedly pessimistic, emphasizing how perceivers "fail" (p. 485), because they "consistently underestimate, and/or make inadequate allowance for, the biasing effects of the questioners' and contestants' roles upon their ability to display general knowledge advantageously" (p. 486). They further claimed that "the phenomenon demonstrated in the . . . present experiment has clear implications for role constrained encounters outside the laboratory" (p. 493).

The unfairness of evaluating the questioner and contestant solely on the basis of their role-constrained performance seems obvious. Yet it may be misleading to regard this "attribution error" in a wholly negative light. The social error of overattribution must, it seems, be preceded by development of a prior and general social *competence*, the ability to recognize and make appropriate attributions for differential task performance in the first place. Overextension of a rule first requires that the rule be acquired; a recognition of how better performance ordinarily implies greater ability must precede any tendency to draw this implication too broadly. A similar developmental sequence is found in many areas of psychology (e.g., language acquisition; see Brown, 1973): Initially, behavior occurs without reference to a rule, then the rule is learned but applied too broadly, then finally the exceptions to the rule are learned. The social roles

effect studied by L. Ross et al. seems to involve a failure to recognize an exceptional (and artificial) case, but also seems to require acquisition and application of the rule itself.

This approach has important implications for how we might interpret differences between individuals who do and do not manifest this attribution error. For example, when someone immune to the social roles effect watches the two individuals interact and then concludes they "look about the same" in ability, is this because he or she is the rare individual able to adjust appropriately for the clever way the experimenter rigged the situation, or does he or she simply fail to infer differential ability from differential performance in the first place? Recently, J. Block and Funder (1986) have obtained relevant data, examining the personality correlates of individual differences in the role effect among 14-year-old subjects. They found that far from being exceptionally sophisticated, those males (and, to a lesser extent, females) who were least vulnerable to the role effect were independently characterized as the least socially competent and least well adjusted. Moreover, individual differences in the role effect correlated positively with self-esteem, and females exhibited the role effect to a significantly greater degree than males, a finding consistent with evidence elsewhere that adolescent females tend to be better attuned socially than males (J. H. Block, 1973; Huston, 1983).

Thus, manifestation of this attribution "error," far from being a symptom of social maladjustment, actually seems associated with a degree of competence, at least in adolescents. To be sure, in Ross et al.'s experiment there was no valid reason for subjects to regard the questioner as any more knowledgeable than the contestant. Yet it also seems likely that in the overwhelming majority of wider, real-life contexts, such differences in apparent performance in fact do reflect differences in ability. Performance may not ordinarily be so constrained by social role, after all, and social roles themselves are not usually conferred quite as arbitrarily as in this experiment. The relation that ordinarily obtains between performance and ability in real life may mean that it is generally more efficient and functional to infer ability directly from performance than to try to adjust for the specific constraints that may be present in every situation encountered.¹¹

The Information Given, and Beyond

In each of the three studies just summarized, the stimulus is an isolated, artificial piece of social reality (cf. Ebbesen & Konecni, 1980). Subjects are presented with an essay purportedly written by a stimulus person, or given a description of someone (as purportedly formulated by "trained psychologists"), or placed into a contrived interpersonal encounter of, to say the least, a highly unusual sort. In each case the subjects' departure from the experimenter's normative expectation takes a specific form: Analogously to the subjects who view Figure 1 in three dimensions, subjects place themselves in, and derive their responses from, a wider context than the partial reality the experimenter has in mind. They base their responses on their knowledge of real-world semantics and real-world correlations, on their perceptions of what the experimenter might be "really" communicating to them or otherwise demanding, or on their application of rules that ordinarily can be counted on to lead to correct conclusions. This wider contextualizing is precisely

what leads to a judgment that is incorrect in terms of the immediate, narrowly defined situation, a response that the experimenter will dub an error and, by frequent implication, a mistake as well.

It seems ironic that going beyond the information given in this way is so often interpreted by social psychologists as symptomatic of flawed judgment. Current thinking in the field of artificial intelligence (AI) is that this propensity is exactly what makes people smarter than computers. Such varied authorities on AI as Dreyfus (1979), Minsky (1975), and Schank and Abelson (1970) agree that the key difficulty in getting computers to understand language or solve natural problems is that computers use only the information they have been explicitly given, rather than bringing in the broad array of real-world knowledge and experience that any human would use. The presuppositions, expectations, and even "biases" that some social psychologists seemingly want to eliminate from the judgment process are the same things that researchers in AI are finding necessary—but extremely difficult—to incorporate into that process.

Recent research suggests that the same principle may apply to human judgments. A few studies have been conducted in which raters were trained to make fewer errors (e.g., to eliminate the halo effect, or the leniency bias), and their accuracy was evaluated by criteria external to the judgment process itself. The results of these studies indicate that such training either fails to improve the relation between judgments and external criteria, or actually makes it worse (see Bernardin & Pence, 1980; Borman, 1975, 1979). Analogously, it is obvious that any successful training program to eliminate the "illusions" in Figures 1 and 2 would cause people to misjudge size and distance in the real world, if not in the lab, and to walk into walls.

Many other frequently demonstrated errors may reflect adaptive processes in the real world. "Illusory correlation" (Chapman, 1967; Chapman & Chapman, 1967; S. T. Fiske, 1980; McArthur & Ginsberg, 1981) can be plausibly reinterpreted as the natural by-product of an adaptive tendency to attend primarily to what is most important (McArthur & Baron, 1983). The tendency to be insensitive to regression effects (e.g., Kahneman & Tversky, 1973; Nisbett & Borgida, 1975; Slovic & Lichtenstein, 1971) may stem from the human observers' adaptive sensitivity to changes in the underlying process (Hogarth, 1981). The "illusion of control" (Langer, 1975) may be useful in preventing depression (cf. Hogarth, 1981; Seligman, 1975). Even the much maligned "fundamental attribution error" will lead to correct judgments in real life to the degree that people are actually consistent in their behavior (cf. Jones, 1979, p. 116). Whether people are consistent or not is a matter of controversy within psychology (Funder, 1983b; West, 1983), but the point is that demonstrations of the error itself are not informative about whether it usually leads to mistakes or to correct judgments in real life (cf. Kenrick & Dantchik, 1983).

¹¹ For the observer subjects in such an experiment, a second consideration is also operative. These observers are literally given no information on which to base their attributions of ability, except the role-constrained performances of actor and observers. As in other experiments with overattribution, the very provision of this information itself may well imply that it is intended to be diagnostic of the judgments to be made and should be used.

Getting a Fix on Accuracy

If research on the process of social judgment does not address the difference between errors and mistakes, we are left to wonder what sort of empirical research can address the accuracy issue: How accurate is social judgment, anyway?

Phrased this way, the question is probably unanswerable. It can only have meaning in relation to a second question: Compared to what? Some authors (e.g., Nisbett & Ross, 1980), have proposed that people are less accurate than they think they are. Others (e.g., Swann, 1984) have suggested that people are more accurate than some psychologists think they are.¹² If one accepts that people are neither infallibly accurate nor wrong all the time, such discussions seem to lose much of their interest, and reduce to an analysis of the difference between a half-full and a half-empty glass (Loftus & Beach, 1982).

However, we should not lose sight of the cumulative effect of the past 10 years of research on error. It is surely no exaggeration to say that the one-sided emphasis on what people do wrong has had a profound effect on many psychologists' and psychology students' views of the accuracy issue. As was proposed at the beginning of this article, the current *Zeitgeist* emphasizes purported flaws in human judgment to the extent that it might well be "news" to assert that people can make global judgments of personality with any accuracy at all.

Such an assertion cannot be assessed by research that examines only *how* judgments are made. Rather, research must let subjects judge real people in real social situations and assess their abilities in relation to the content of the judgments themselves: In a realistic situation, are the judgments right or wrong?

In principle, there are three ways to follow this advice, not all equally practical. The simplest approach might be to note the general success with which people manage to negotiate their social worlds. Gibson (1979) describes analogously how he decided that visual illusions tend to underestimate human perceptual abilities when he began studying how pilots manage simply to look out their windows and then land their airplanes. In the realm of person perception, we could note how people by and large do manage to negotiate their social worlds, choose friends, complete transactions, and perform jobs. Unfortunately, it is also true that people sometimes misinterpret their friends' behavior, hire the wrong job candidates, and enrich astrologers (see Nisbett & Ross, 1980, p. 252). Thus, an overall evaluation of real-world social reasoning tends to degenerate again into a distinction between half-full and half-empty glasses.

A second, more complex possibility might be research that presents subjects with social stimuli in the laboratory, that although artificial were to be fairly "representative" of social stimuli in real life (Brunswick, 1956; Hammond, Hamm, Grasia, & Pearson, 1984). Any combination of information that cannot be found in social reality would be prohibited in the lab. Instead, stimuli would be used in the combinations that are most common and important in real life. This is exactly the sort of research that could distinguish precisely between errors and mistakes, by evaluating when realistic information is processed in such a way as to produce incorrect outcomes.

Such a research program would be difficult and complex. The essential first step should be an inventory of the sources of information that actually exist in the interpersonal world, and an assessment of the frequency, informativeness, and interre-

lations of each of these sources. The second step should be to use this inventory to construct artificial arrangements of social stimuli that are faithful representations of important situations found in real life. The final step should be to study how subjects process and integrate such realistic arrays of information. Somewhat discouragingly, perceptual psychologists have already found this sort of "representative" research program extremely difficult to carry out even in the much simpler domain of vision (Brunswick, 1956; Hammond, 1966). A few psychologists interested in general, nonsocial judgments have nonetheless taken this route (e.g., Hogarth & Makridakis, 1981; Kleinmütz & Kleinmütz, 1981), and social psychologists have just begun (McArthur & Baron, 1983; Nisbett, Krantz, Jepson, & Kunda, 1983). As this line of research continues to develop, representative research may eventually offer a promising method for studying the accuracy issue.

A third, final, and more immediate possibility is research that goes directly to the heart of the accuracy issue by examining, according to various external criteria, whether real-life judgments are right or wrong. The criteria can be realistic and pragmatic, if not ultimate.

Consider how laypersons regard "accuracy" in social judgment. In realistic contexts, accuracy has two meanings. The first is agreement. In ordinary discourse it is assumed that two observers' judgments, to both be accurate, must generally agree with one another. The beginning of any conversation about another person, for example, frequently begins with some verbal preliminaries to assess whether the impressions held by both individuals generally agree. If it turns out that the observers have markedly different impressions, suspicion will arise that at least one of them must be wrong, and the subsequent course of conversation is likely to be an attempt to negotiate an impression that can be agreed on.

A second pragmatic definition of accuracy is the ability to predict behavior. A prospective employer evaluating an applicant's "conscientiousness" is less interested in the deep structure of his or her personality than in whether he or she will usually appear at work on time and carry out assigned duties diligently. A prospective friend evaluating someone's "sociability," or a banker evaluating someone's "honesty," is attempting to predict future behavior, not settle ultimate questions of individual identity (cf. Swann, 1984).

This perspective yields an insight into how psychologists might realistically begin to evaluate the accuracy of social judg-

¹² Swann's (1984) suggestion is based on the distinction between what he refers to as subjects' "circumscribed" accuracy, within their limited social worlds, and their more "global" accuracy, that is, accuracy in relation to behaviors they have not seen or contexts in which they have not observed the object of judgment. "To the extent that perceivers aspire to circumscribed rather than to global accuracy, they may enjoy higher levels of accuracy than past research [on error] suggests" (p. 462). A given subject's accuracy within a relatively narrow domain, "close to home," probably is better than his or her accuracy at large. However, the thesis here is that traditional process-oriented research on error is not informative about either circumscribed or global accuracy. Moreover, the criteria for accuracy promoted later in the article, which include agreement between different judges with necessarily different observational experience, and the ability of their judgments to predict behavior in a unique setting, seem more relevant to what Swann would call "global" rather than "circumscribed" accuracy.

ments: Do they agree with each other, and do they predict behavior? As general propositions, the answer to both of these questions is "yes."

Some Evidence on the Accuracy of Judgments

Correlations Between Self-Judgments and Others' Judgments

When agreement between two different judges, such as the self and another person, is found, the judgments that agree might be either right or wrong. For example, some degree of agreement could arise artifactually if the rater and ratee have ever discussed and negotiated what they both think the ratee is like (McClelland, 1972). On the other hand, recent research suggests that when people's verbal self-presentations contradict their behavior, raters will ignore the words in favor of the deeds (Amabile & Kabat, 1982). Interpretation is more certain when two judgments fail to agree: At least one of them must be wrong. Some degree of agreement between judges is necessary if not sufficient for accuracy, therefore, and so research on such agreement is relevant to the accuracy issue.

Although, remarkably, articles continue to appear claiming that agreement between self-judgments of personality and judgments by others is low (e.g., Kammann, Smith, Martin, & McQueen, 1984; Shrauger & Schoeneman, 1979), in fact such findings obtain only when the peer raters do not know their subjects well or unreliable rating scales are used. Otherwise, such correlations tend to be on the order of .30 to .60 (see, e.g., Cheek, 1982; also Andersen, 1984; Edwards & Klockars, 1981; Funder, 1980; Funder & Dobroth, in press; Goldberg, Norman, & Schwartz, 1980; Hase & Goldberg, 1967; McCrae, 1982; Monson, Tanke, & Lund, 1980; Paunonen & Jackson, 1985; Woodruffe, 1985). Moreover, at least four different studies have independently verified the additional fact that such correlations are higher when the person in question describes him- or herself as consistent on the trait being evaluated, versus as variable on that dimension (Bem & Allen, 1974; J. B. Campbell, 1985; Cheek, 1982; Kenrick & Stringfield, 1980). Other studies establish that as the judge gets to know his or her target better, agreement between his or her judgments and the target's own self-judgments steadily improves (e.g., Jackson, Neill, & Bevan, 1973; Kusyszyn, 1968; Norman & Goldberg, 1966). Thus, if proper care is taken to use valid rating scales and peer judges who know us well, then by and large we do tend to "see ourselves as others see us" (Funder, 1980; Norman, 1969).

Correlations Among Others' Judgments

Some of the studies on self-other agreement just cited also demonstrate substantial correlations among different other persons' judgments of someone's personality (see also D. W. Fiske & Cox, 1960; Hollander, 1957; Kane & Lawler, 1978). Previously unpublished data are available that speak directly to this point. Each of 37 Stanford University undergraduates was assessed by two peers using the Q-sort method (J. Block, 1961/1978) of personality description. About 2 years later, similar data were gathered on a second sample of 69 undergraduates. (These data were gathered in connection with studies on other topics by Bem & Funder, 1978, and Funder, 1982a.) In each

sample, more than half of the 100 personality items in the Q-set attained significant interjudge agreement (at $p < .05$; specifically, 55 items in Sample I and 51 items in Sample II). Thirty items attained significant interjudge convergence independently in both samples.¹³ This set of replicated items appears in Table 1. Table 1 reveals that substantial agreement between judgments can repeatedly be found on a wide variety of dimensions, including some that seem relatively global and even inferential in nature (see also Funder & Dobroth, in press).

Correlations Between Behavior and Judgments of Personality

This third type of data is perhaps the most germane of all to the accuracy issue. Interestingly, when the great perceptual psychologist Egon Brunswik (1944) decided that arbitrary stimuli in the lab were inadequate tools for assessing the adaptive mechanisms of vision, he moved his own research into the real world. Rather than exclusively study visual illusions, he followed a subject around in the environment and asked her to estimate the distances of various objects. Brunswik found that her judgments of distance correlated with actual distance (after logarithmic transformation) with an $r = .99$.

In the domain of person perception we probably cannot hope for correlations of .99, but we can follow Brunswik's example and calculate the correlations between subjects' judgments of personality and what the persons being judged actually do. This advice has been followed all along by a number of industrial psychologists, who did not give up on accuracy research even after Cronbach (1955). The pragmatic interest of these investigators in the best ways to predict future job performance continued to nourish a lively literature on the relation between peer judgments and future behavioral performance. Most of this research has appeared in the pages of such journals as *Organizational Behavior and Human Performance* and the *Journal of Applied Psychology* (for reviews see Kane & Lawler, 1978; Lewin & Zwany, 1976). Ratings by peers have been used successfully to predict such diverse criteria as performance as a Naval officer (Hollander, 1965), life insurance sales (measured in dollars, Mayfield, 1972), and sales of data processing systems (Waters & Waters, 1970). In general, peer ratings have become a well-respected means for predicting the performance of individuals in industrial and organizational settings.

Naturally, these investigators have been less interested in how well ratings of general personality traits predict behavior not relevant to industrial performance, but recent research by more theoretically inclined psychologists provides several examples. In various studies "informants" who knew the subjects well provided descriptions of subjects' personalities based on their judgment from real-life acquaintance, and operationalized through the Q-sort method (J. Block, 1961/1978). These same subjects were then placed in experimental situations in which their actual overt behavior could be directly measured. Typically, the correlations between the personality ratings and overt

¹³ Because the distinction between judges in each pair is arbitrary, agreement was assessed using intraclass correlations (Rosenthal & Rosnow, 1984), which are interpreted in exactly the same way as a Pearson r , and are tested for significance with a one-tailed test.

Table 1
Interobserver Agreement in Personality Judgments

Q-Item	Intraclass correlations	
	Sample I (n = 37)	Sample II (n = 69)
90. Concerned with philosophical problems	.52	.53
80. Interested in opposite sex	.59	.43
52. Behaves in an assertive fashion	.36	.55
81. Is physically attractive	.51	.39
62. Rebellious and nonconforming	.57	.30
31. Regards self physically attractive	.56	.35
91. Power oriented; values power in self and others	.52	.39
8. High intellectual capacity	.53	.31
99. Is self-dramatizing (exaggerates emotion)	.43	.41
25. Tends toward overcontrol of needs and impulses	.49	.32
84. Is cheerful	.35	.45
37. Is guileful and deceitful	.44	.35
28. Tends to arouse liking and acceptance	.45	.31
4. Is a talkative individual	.39	.36
65. Pushes and stretches limits	.42	.31
21. Arouses nurturant feelings in others	.44	.29
40. Vulnerable to threat; fearful	.41	.30
97. Emotionally bland	.43	.28
54. Emphasizes being with others; gregarious	.33	.35
7. Favors conservative values	.36	.30
71. High aspiration level for self	.33	.33
74. Feels satisfied with self	.31	.33
29. Is turned to for advice and reassurance	.36	.27
92. Has social poise and presence	.32	.30
55. Is self-defeating	.32	.28
67. Is self-indulgent	.28	.31
27. Shows condescending behavior	.36	.23
66. Enjoys aesthetic impressions (e.g., art)	.33	.23
86. Repressive or disassociative tendencies	.30	.25
93. Sex-typed (masculine or feminine)	.28	.20

Note. Item wordings in this table are abbreviated. All correlations in this table are significant at $p < .05$ (one-tailed) or better.

behavior have been many, significant, meaningful, and sometimes quite large.

A topic of repeated research has been delay of gratification. At least three separate studies have placed children into situations in which their resistance to temptation was measured directly. For example, Bem and Funder (1978) measured in minutes how long 3- to 4-year-olds were willing to wait for a preferred snack reward. Soon after, these investigators asked the parents of these children to describe their children's personalities, using the Child Q-sort. Funder, Block, and Block (1983) measured the delay of gratification behavior of 4-year-olds and obtained personality judgments from their school teachers when the children were still 4 and later when they were 7 and 11 years of age. Mischel (1984) measured delay in preschoolers but did not obtain personality judgments from their parents until the subjects were (on average) more than 16-years old. In all three cases, theoretically meaningful and powerful patterns of correlations were found between delay behavior and judgments of personality. In each experiment, more than twice as many correlates appeared than chance would predict, even though (a) the exact experimental situations in which delay behavior was measured did not resemble any real-life situations in which the informants could have seen the subjects and (b) in the latter two studies several years had elapsed between the behavior and the personality judgment.

Numerous other examples of the capacity of lay judgments to predict behavior could be cited, from a variety of behavioral domains (e.g., Funder, 1983a; Funder & Harris, 1986b). Moskowitz and Schwarz (1982) have shown that it takes 1,440 direct measurements of behavior, made over a span of 8 weeks, to yield a score that approaches the reliability and validity of the average personality judgments of just three informants. Predictive validity of this sort could simply not exist unless there were a substantial degree of "global accuracy" to observers' judgments of personality, the various errors to which they are prone notwithstanding.

Size of the Relations

Still, many psychologists would not regard the data just summarized as terribly impressive. Although perhaps granting that interrater agreement and the predictability of behavior from judgments of personality can go beyond mere chance, they would also note that the actual correlation coefficients usually lie in the range of .30 to .40 (although they occasionally go higher). The usual practice with such correlations is to square them and thereby compute that somewhere between 9% and 16% of the variance has been "accounted for." Such phrases, although not completely self-explanatory (O'Grady, 1982), do sound unimpressive. To remark disparagingly on the small size

Table 2
*Prediction of a Dichotomous Outcome Based on
 a Correlation of $r = .40$*

Score on predictive variable	Outcome		Marginals
	Success	Failure	
High	70	30	100
Low	30	70	100
Marginals	100	100	200

of "personality coefficients" such as these, therefore, has nearly attained the status of a cliché in the psychological literature.

Three recent developments suggest that it may be time to reassess this conventional viewpoint. Empirically, Funder and Ozer (1983) found that correlations within the range of .30 to .40 characterize the strength of some of the most important effects in experimental social psychology; in particular the effects of incentive on attitude change (Festinger & Carlsmith, 1959), of "hurry" and number of onlookers on bystander intervention (Darley & Batson, 1973; Darley & Latane, 1968), and of proximity of victim and authority on obedience (Milgram, 1975). No one seems to doubt that these are important effects; yet their sizes account for small proportions of the variance by the usual calculation (see also Funder & Harris, 1986a).

Does this analysis merely reveal that these other effects, too, are embarrassingly small? Rosenthal and Rubin (1979, 1982) suggest that the answer is "no." These investigators have invented a method for displaying the size of effects clearly and vividly, a method they call the binomial effect size display. Consider, for example, a dichotomized prediction task in which a judgment (high or low) of a personality variable is used to predict success or failure at some important task in life. Assume that the correlation between judgment and outcome is merely .40, a plausible figure given the results summarized previously. The predictive result is shown in Table 2. The binomial effect size display reveals results that are more encouraging than some psychologists seem to be aware; specifically, that a correlation that explains only 16% of the variance also can yield predictions that are 70% accurate.¹⁴

A third consideration is the recent analysis by Ozer (1985), which suggests that the widely practiced method for computing percentage of variance explained may often be inappropriate in the first place (see also Abelson, 1985; D'Andrade & Dart, 1983). Under the circumstances that obtain in most cases in which "personality coefficients" are calculated (where both predictor and outcome are representatives of the same unmeasured latent variable), the percentage of shared variance is simply the unsquared correlation coefficient itself. For example, if a judgment of personality correlates $r = .40$ with a concrete behavioral measure, then the underlying trait "accounts for" 40% of the total variance. This recognition is actually not new; Ozer's historical survey traces it back more than 50 years (Tryon, 1929). Yet its rediscovery may have the effect of rendering the contemporary method of calculating "variance explained" obsolete.

The data currently available are sufficient to confer some respectability on our judgments of personality, therefore. They often demonstrate convergent and predictive validity to a de-

gree that descriptions such as "better than chance" do not do justice. However, this conclusion, although important, is merely preliminary to the ultimately more interesting question of *when* judgments are more and less likely to be accurate. This question cannot be addressed by studying error in the laboratory, but only by research that takes the difficult step of constructing criteria for accuracy in real life.

Remaining Questions

This article has been highly critical of the currently dominant paradigm that studies error in social judgment. Errors are not necessarily mistakes, I have argued, because judgments (and judgment processes) that are incorrect in terms of a limited experimental context may be correct when applied to a wider, more realistic context—a context, in other words, in which judgments are usually made and have real-life consequences. Or, of course, they also may not be correct in a wider context. Yet research on error cannot settle the question. For that reason it is not relevant to the accuracy of human judgment in real life.

Moreover, when judgments do follow normative strategies and yield correct judgments of experimental stimuli, they cannot be assumed to reflect processes that lead to correct judgments in the real world. This latter inference is hardly ever drawn, of course. However, it is worth bearing in mind that research that defines error in terms of deviation from normative models is not relevant to accuracy in either the positive or negative direction. Normative models of judgment have a limited range of application; they cannot be relied on to produce correct judgments anywhere outside specific experimental contexts (Hogarth, 1981; Kruglanski, Friedland, & Farkash, 1984; Swann, 1984; Trope, Bassok, & Alon, 1984). That is why, as Kruglanski and Ajzen (1983) write, "searching for the paragon of rationality may well have a 'will of the wisp' quality" (p. 37).

Still, the study of error, in social judgment and in judgment generally, is an important and valuable part of psychology. Along with the study of reaction time, error analysis is one of the two most useful tools for learning about how people think. Ironically, however, its great usefulness for studying process is not what has led this research to be so extraordinarily popular. Rather, its popularity is largely if not exclusively due to its apparent and, it turns out, illusory implications for accuracy. That is why the first half of this article had to be so negative. Only when psychologists understand what error research *cannot* deliver, will they fully appreciate what it can.

Another issue needs clarification. Some readers of the present analysis have commented that it has, overall, an almost Panglossian tone, specifically, that it seems to assume people are seldom wrong and that most if not all errors reflect adaptive processes. One colleague commented, "You assume the human judge is innocent until *proven* guilty."

There is some truth to this comment. The second part of this article *is* optimistic in contrast to the usual view, and presents evidence that people agree better and make better predictions

¹⁴ The computational method is to define the marginals as shown, then multiply the correlation by 100 (to remove the decimal), divide by 2, add 50, and insert this figure in the upper left-hand cell. Because this 2×2 table has 1 degree of freedom, the other cell entries can be obtained by subtraction (Rosenthal & Rubin, 1982).

than some psychologists seem to think. However, it should be emphasized that as mentioned previously, the analysis of error research that precedes this evidence is itself completely neutral on the accuracy issue. It does not assume that all errors reflect adaptive processes, but asserts only that some might. The analysis does claim that people are not necessarily wrong when they make errors in the laboratory, but it should be taken equally to imply that people are not necessarily right when they do not make errors.

That observation brings us back to the central point of this article: The criteria for evaluating social judgments reside not in the lab, where all you can study is the process, but in the world, where their content is. The present article has proposed that the two most immediately useful candidates for real-world criteria are the agreement between the judgments of different observers and the ability of judgments to predict behavior, and has summarized research that demonstrates an impressive amount of both. Yet as usual when a new line of research begins to open (or reopen), those findings raise more questions than they answer.

A fundamental question involves how to apply the criteria for accuracy in future research. The answer will not always be straightforward. When a personality judgment fails to predict a behavior, for example, this cannot necessarily be taken to mean the judgment was wrong. Maybe the wrong behavior was measured. No single, simple set of criteria will suffice for evaluating the accuracy of social judgments. Rather, the process of validating a judgment must be quite like the process of "construct validation" used to evaluate other sorts of psychological measurement (cf. Cronbach & Meehl, 1955; Loewinger, 1957). This process is complex and involves the analysis of a variety of criteria and their interrelations. Construct validation is an important and long-standing part of the technology of personality assessment (see Wiggins, 1973, chap. 9), and its incorporation into the evaluation of interpersonal judgment seems likely to develop an important intersection between the traditional concerns of personality and social psychology.

Other, more specific questions arise from recent research. What are the implications of research on conformity (which can produce a consensus between judges) and the use of agreement as a criterion? Can the theory and methodology used in studying the attitude-behavior relation (e.g., Fishbein & Ajzen, 1974) be applied to the use of behavioral prediction as a criterion for accuracy? Hampson, John, and Goldberg (1986) suggest that an important difference between traits is that some are relatively broad and others are more narrow. As in attitude-behavior relations, will judgments of narrow traits best predict relatively narrow behavioral criteria, whereas broad traits better suit the prediction of large, diverse patterns of behavior (cf. Ajzen, 1982)? Another question is raised by the recent finding that subjects agree with each other best when rating individuals on those traits that seem, according to other subjects, relatively "easy to judge" (Funder & Dobroth, in press). Will these "easy" traits also yield higher correlations with behavioral criteria?

Finally, some questions have a long history, but have been neglected in recent years. What sorts of tools for judgment tend to maximize interrater agreement and relations between judgments and behavior? What dimensions of personality are most likely to be judged accurately? What sort of experience with the person to be judged—structured, unstructured, intimate, or

detached—best promotes accuracy in subsequent judgments? How much of an acquaintance is necessary for accurate judgment—is there a bare minimum, or a point of diminishing return? Can the judgment of personality be improved by instruction or experience—if so, what kinds?

Questions like these once interested psychologists like Allport (1937), Guilford (1936), and Murray (1938; cf. Ozer, 1979), but apparently were forgotten. It may finally be time to pick up where these investigators left off. If one is interested in the accuracy of social judgments, the relevant empirical questions have nothing to do with error. Rather, they are (a) Do the judgments agree with each other? (b) Do they predict behavior?

References

- Abelson, R. (1985). A variance explanation paradox: When a little is a lot. *Psychological Bulletin*, *97*, 129–133.
- Adler, J. E. (1984). Abstraction is uncooperative. *Journal for the Theory of Social Behaviour*, *14*, 165–181.
- Ajzen, I. (1982). On behaving in accordance with one's attitudes. In M. P. Zanna, E. T. Higgins, & C. P. Herman (Eds.), *Consistency in social behavior: The Ontario Symposium* (pp. 3–16). Hillsdale, NJ: Erlbaum.
- Allport, G. W. (1937). *Personality: A psychological interpretation*. New York: Holt, Rinehart & Winston.
- Amabile, T. M., & Kabat, L. G. (1982). When self-descriptions contradict behavior: Actions do speak louder than words. *Social Cognition*, *1*, 311–325.
- Andersen, S. (1984). Self-knowledge and social inference: II. The diagnosticity of cognitive/affective and behavioral data. *Journal of Personality and Social Psychology*, *46*, 294–307.
- Asch, S. E. (1946). Forming impressions of personality. *Journal of Abnormal and Social Psychology*, *41*, 258–290.
- Asch, S., & Zukier, H. (1984). Thinking about persons. *Journal of Personality and Social Psychology*, *46*, 1230–1240.
- Bem, D. J., & Allen, A. (1974). On predicting some of the people some of the time: The search for cross-situational consistencies in behavior. *Psychological Review*, *81*, 506–520.
- Bem, D. J., & Funder, D. C. (1978). Predicting more of the people more of the time: Assessing the personality of situations. *Psychological Review*, *85*, 485–501.
- Berman, J. S., Read, S. J., & Kenny, D. A. (1983). Processing inconsistent social information. *Journal of Personality and Social Psychology*, *45*, 1211–1224.
- Bernardin, H. J., & Pence, E. C. (1980). Effects of rater training: Creating new response sets and decreasing accuracy. *Journal of Applied Psychology*, *65*, 60–66.
- Bernstein, W. M., & Davis, M. H. (1982). Perspective taking, self-consciousness, and accuracy in person perception. *Basic and Applied Social Psychology*, *3*, 1–19.
- Block, J. (1977). Correlational bias in observer ratings: Another perspective on the Berman and Kenny study. *Journal of Personality and Social Psychology*, *35*, 873–880.
- Block, J. (1978). *The Q-sort method in personality assessment and psychiatric research*. Palo Alto, CA: Consulting Psychologists Press. (Original work published 1961)
- Block, J., & Funder, D. C. (1986). Social roles and social perception: Individual differences in attribution and "error." *Journal of Personality and Social Psychology*, *51*, 1200–1207.
- Block, J. H. (1973). Conceptions of sex role: Some cross-cultural and longitudinal perspectives. *American Psychologist*, *28*, 512–526.
- Borman, W. C. (1975). Effects of instructions to avoid halo error on reliability and validity of performance ratings. *Journal of Applied Psychology*, *62*, 64–69.

- Borman, W. C. (1979). Format and training effects on rating accuracy and rater errors. *Journal of Applied Psychology*, 64, 410-421.
- Brown, R. (1973). *A first language: The early stages*. Cambridge, MA: Harvard University Press.
- Brunswik, E. (1944). Distal focussing of perception: Size-constancy in a representative sample of situations. *Psychological Monographs*, 56 (1, Whole No. 254).
- Brunswik, E. (1956). *Perception and the representative design of experiments*. Berkeley: University of California Press.
- Campbell, D. T. (1959). Systematic error on the part of human links in communication systems. *Information and Control*, 1, 334-369.
- Campbell, J. B. (1985, April). *Cross-situational consistency of personality attributes*. Paper presented at the Annual Meetings of the Eastern Psychological Association, Boston.
- Chapman, L. J. (1967). Illusory correlation in observational report. *Journal of Verbal Learning and Verbal Behavior*, 6, 151-155.
- Chapman, L. J., & Chapman, J. P. (1967). Genesis of popular but erroneous psychodiagnostic observations. *Journal of Abnormal Psychology*, 72, 193-204.
- 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-Szalanski, J. J. J., & Beach, L. R. (1984). The citation bias: Fad and fashion in the judgment and decision literature [Comment]. *American Psychologist*, 39, 75-78.
- Clark, H. H., & Clark, E. V. (1977). *Psychology and Language*. New York: Harcourt Brace Jovanovich.
- Cohen, L. J. (1981). Can human irrationality be experimentally demonstrated? *Behavioral and Brain Sciences*, 4, 317-331.
- Cook, M. (1984). *Issues in person perception*. London: Methuen.
- Crandall, C. S. (1984). The overcitation of examples of poor performance: Fad, fashion, or fun [Comment]? *American Psychologist*, 39, 1499-1500.
- Cronbach, L. J. (1955). Processes affecting scores on "understanding of others" and "assumed similarity." *Psychological Bulletin*, 52, 177-193.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281-302.
- D'Andrade, R. G., & Dart, J. (1983). *The interpretation of τ versus τ^2* . Unpublished manuscript, University of California at San Diego.
- Darley, J. M., & Batson, C. D. (1973). "From Jerusalem to Jericho": A study of situational and dispositional variables in helping behavior. *Journal of Personality and Social Psychology*, 27, 100-108.
- Darley, J. M., & Latane, B. (1968). Bystander intervention in emergencies: Diffusion of responsibility. *Journal of Personality and Social Psychology*, 8, 377-383.
- Dember, W. N., & Warm, J. S. (1979). *Psychology of perception* (2nd ed.). New York: Holt, Rinehart & Winston.
- DePaulo, B. M. (1978). Accuracy in predicting situational variations in help-seekers' responses. *Personality and Social Psychology Bulletin*, 4, 330-333.
- Dreyfus, H. L. (1979). *What computers can't do: The limits of artificial intelligence*. New York: Harper & Row.
- Dymond, R. F. (1949). A scale for the measurement of empathic ability. *Journal of Consulting Psychology*, 13, 127-133.
- Dymond, R. F. (1950). Personality and empathy. *Journal of Consulting Psychology*, 14, 343-350.
- Ebbesen, E. B., & Konecni, V. J. (1980). On the external validity of decision-making research: What do we know about decisions in the real world? In T. S. Wallsten (Ed.), *Cognitive processes in choice and decision behavior* (pp. 21-45). Hillsdale, NJ: Erlbaum.
- Edwards, A. L., & Klockars, A. J. (1981). Significant others and self-evaluation: Relationships between perceived and actual evaluations. *Personality and Social Psychology Bulletin*, 7, 244-251.
- Einhorn, H. J., & Hogarth, R. M. (1985). Ambiguity and uncertainty in probabilistic inference. *Psychological Bulletin*, 92, 433-461.
- Estes, S. G. (1938). Judging personality from expressive behavior. *Journal of Abnormal and Social Psychology*, 33, 217-236.
- Evans, J. St. B. T. (1984). In defense of the citation bias in the judgment literature [Comment]. *American Psychologist*, 39, 1500-1501.
- Festinger, L., & Carlsmith, J. M. (1959). Cognitive consequences of forced compliance. *Journal of Abnormal and Social Psychology*, 58, 203-210.
- Fishbein, M., & Ajzen, I. (1974). Attitudes toward objects as predictors of single and multiple behavioral criteria. *Psychological Review*, 81, 59-74.
- Fiske, D. W., & Cox, J. A. (1960). The consistency of rating by peers. *Journal of Applied Psychology*, 44, 11-17.
- Fiske, S. T. (1980). Attention and weight in person perception: The impact of negative and extreme behaviors. *Journal of Personality and Social Psychology*, 38, 889-908.
- Fiske, S. T., & Taylor, S. E. (1984). *Social cognition*. New York: Random House.
- 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. (1982a). On assessing social psychological theories through the study of individual differences: Template matching and forced compliance. *Journal of Personality and Social Psychology*, 43, 100-110.
- Funder, D. C. (1982b). On the accuracy of dispositional vs. situational attributions. *Social Cognition*, 1, 205-222.
- Funder, D. C. (1983a). The "consistency" controversy and the accuracy of personality judgments. *Journal of Personality*, 51, 346-359.
- Funder, D. C. (1983b). Three issues in predicting more of the people: A reply to Mischel and Peake. *Psychological Review*, 90, 283-289.
- Funder, D. C., Block, J. H., & Block, J. (1983). Delay of gratification: Some longitudinal personality correlates. *Journal of Personality and Social Psychology*, 44, 1198-1213.
- Funder, D. C., & Dobroth, K. M. (in press). Differences between traits: Properties associated with inter-judge agreement. *Journal of Personality and Social Psychology*.
- Funder, D. C., & Harris, M. J. (1986a). Experimental effects and person effects in delay of gratification [Comment]. *American Psychologist*, 41, 476-477.
- Funder, D. C., & Harris, M. J. (1986b). On the several facets of personality assessment: The case of social acuity. *Journal of Personality*, 54, 528-550.
- Funder, D. C., & Ozer, D. J. (1983). Behavior as a function of the situation. *Journal of Personality and Social Psychology*, 44, 107-112.
- Funk & Wagnall's standard college dictionary: Text edition*. (1963). New York: Harcourt, Brace & World.
- Gage, N. L., & Cronbach, L. J. (1955). Conceptual and methodological problems in interpersonal perception. *Psychological Review*, 62, 411-422.
- Gettys, C. F., Kelly, C. W., III, & Peterson, C. R. (1973). The best guess hypothesis in multistage inference. *Organizational Behavior and Human Performance*, 10, 364-373.
- Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.
- Goffman, E. (1974). *Frame analysis*. New York: Harper & Row.
- Goldberg, L. R., Norman, W. T., & Schwartz, E. (1980). The comparative validity of questionnaire data (16PF scales) and objective test data (O-A battery) in predicting five peer-rating criteria. *Applied Psychological Measurement*, 4, 183-194.
- Gregory, R. L. (1971). Visual illusions. In R. C. Atkinson (Ed.), *Contemporary psychology* (pp. 167-177). San Francisco: Freeman.
- Grice, H. P. (1975). Logic in conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics* (Vol. 3, pp. 41-58). New York: Academic Press.
- Guilford, J. P. (1936). *Psychometric methods*. New York: McGraw-Hill.
- Hammond, K. R. (1966). Probabilistic functionalism: Egon Brunswik's

- integration of the history, theory, and method of psychology. In K. R. Hammond (Ed.), *The psychology of Egon Brunswik* (pp. 15–80). New York: Holt, Rinehart & Winston.
- Hammond, K. R., Hamm, R. M., Grassia, J., & Pearson, T. (1984). *The relative efficacy of intuitive and analytical cognition: A second direct comparison*. Boulder, CO: Center for Research on Judgment and Policy.
- Hampson, S. E., John, O. P., & Goldberg, L. R. (1986). Category breadth and hierarchical structure in personality: Studies of asymmetries in judgments of trait implications. *Journal of Personality and Social Psychology*, *51*, 37–54.
- Harackiewicz, J. M., & DePaulo, B. M. (1982). Accuracy of person perception: A component analysis according to Cronbach. *Personality and Social Psychology Bulletin*, *8*, 247–256.
- Harvey, J. H., Town, J. P., & Yarkin, K. L. (1981). How fundamental is "the fundamental attribution error"? *Journal of Personality and Social Psychology*, *43*, 345–346.
- Hase, H. D., & Goldberg, L. R. (1967). Comparative validity of different strategies of constructing personality inventory scales. *Psychological Bulletin*, *67*, 231–248.
- Hastie, R., Ostrom, T. M., Ebbesen, E. B., Wyer, R. S., Hamilton, D. L., & Carlston, D. E. (1980). *Person memory: The cognitive basis of social perception*. Hillsdale, NJ: Erlbaum.
- Hastorf, A. H., & Bender, I. E. (1952). A caution respecting the measurement of empathic ability. *Journal of Abnormal and Social Psychology*, *47*, 574–576.
- Heider, F. (1958). *The psychology of interpersonal relations*. New York: Wiley.
- Helmholtz, H. von (1903). *Popular lectures on scientific subjects* (E. Atkinson, Trans.). New York: Green. (Original work published 1881)
- Hogarth, R. M. (1981). Beyond discrete biases: Functional and dysfunctional aspects of judgmental heuristics. *Psychological Bulletin*, *90*, 197–217.
- Hogarth, R. M., & Makridakis, S. (1981). The value of decision making in a complex environment: An experimental approach. *Management Science*, *27*, 93–107.
- Hollander, E. P. (1957). The reliability of peer nominations under various conditions of administration. *Journal of Applied Social Psychology*, *41*, 85–90.
- Hollander, E. P. (1965). Validity of peer nominations in predicting a distant performance criterion. *Journal of Applied Social Psychology*, *49*, 434–438.
- Huston, A. C. (1983). Sex-typing. In E. M. Hetherington (Ed.), *Handbook of child psychology: Vol. 4. Socialization, personality and social development* (pp. 387–467). New York: Wiley.
- Jackson, D. N. (1982). Some preconditions for valid person perception. In M. P. Zanna, E. T. Higgins, & C. P. Herman (Eds.), *Consistency in social behavior: The Ontario Symposium* (pp. 251–279). Hillsdale, NJ: Erlbaum.
- Jackson, D. N., Neill, J. A., & Bevan, A. R. (1973). An evaluation of forced-choice and true–false item formats in personality assessment. *Journal of Research in Personality*, *7*, 21–30.
- Jones, E. E. (1979). The rocky road from acts to dispositions. *American Psychologist*, *34*, 107–117.
- Jones, E. E. (1985). Major developments in social psychology during the past five decades. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology* (3rd ed., Vol. 1, pp. 47–107). New York: Random House.
- Jones, E. E., & Davis, K. E. (1965). From acts to dispositions: The attribution process in person perception. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 2, pp. 219–266). New York: Academic Press.
- Jones, E. E., & Harris, V. A. (1967). The attribution of attitudes. *Journal of Experimental Social Psychology*, *3*, 1–24.
- Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. New York: Cambridge University Press.
- Kahneman, D. T., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, *80*, 237–251.
- Kammann, R., Smith, R., Martin, C., & McQueen, M. (1984). Low accuracy in judgments of others' psychological well-being as seen from a phenomenological perspective. *Journal of Personality*, *52*, 107–123.
- Kane, J. S., & Lawler, E. E., III. (1978). Methods of peer assessment. *Psychological Bulletin*, *85*, 555–586.
- Kelley, H. H. (1967). Attribution theory in social psychology. In D. Levine (Ed.), *Nebraska Symposium on Motivation* (Vol. 15, pp. 192–240). Lincoln: University of Nebraska Press.
- Kelley, H. H. (1973). The process of causal attribution. *American Psychologist*, *28*, 107–128.
- Kenrick, D. T., & Dantchik, A. (1983). Interactionism, idiographics, and the social psychological invasion of personality. *Journal of Personality*, *51*, 286–307.
- 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.
- Kleinmutz, D. N., & Kleinmutz, B. (1981). Decision strategies in simulated environments. *Behavioral Science*, *26*, 294–305.
- Kruglanski, A. W., & Ajzen, I. (1983). Bias and error in human judgment. *European Journal of Social Psychology*, *13*, 1–44.
- Kruglanski, A. W., Friedland, N., & Farkash, E. (1984). Lay persons' sensitivity to statistical information: The case of high perceived applicability. *Journal of Personality and Social Psychology*, *46*, 503–518.
- Kusyszyn, I. (1968). A comparison of judgmental methods with endorsements in the assessment of personality traits. *Journal of Applied Psychology*, *52*, 227–253.
- Langer, E. J. (1975). The illusion of control. *Journal of Personality and Social Psychology*, *32*, 311–328.
- Lanning, K. (1985). *Normative decisions need be neither rational nor optimal: Some psychological considerations in the study of decision-making*. Unpublished manuscript, Institute of Personality Assessment and Research, University of California at Berkeley.
- Lewin, A. Y., & Zwany, A. (1976). Peer nominations: A model, literature critique and a paradigm for research. *Personnel Psychology*, *29*, 423–447.
- Lindblom, C. E. (1959). The science of "muddling through." *Public Administration Review*, *19*, 79–88.
- Lindsay, P. H., & Norman, D. A. (1972). *Human information processing*. New York: Academic Press.
- Loevinger, J. (1957). Objective tests as instruments of psychological theory [Monograph]. *Psychological Reports*, *3*, 635–694.
- Loftus, E. F., & Beach, L. R. (1982). Human inference and judgment: Is the glass half empty or half full? *Stanford Law Review*, *34*, 939–956.
- Mayfield, E. C. (1972). Value of peer nominations in predicting life insurance sales performance. *Journal of Applied Psychology*, *56*, 319–323.
- McArthur, L. Z. (1972). The how and what of why: Some determinants and consequences of causal attribution. *Journal of Personality and Social Psychology*, *22*, 171–193.
- McArthur, L. Z. (1976). The lesser influence of consensus than distinctiveness information on causal attributions: A test of the person–thing hypothesis. *Journal of Personality and Social Psychology*, *33*, 733–742.
- McArthur, L. Z., & Baron, R. M. (1983). Toward an ecological theory of social perception. *Psychological Review*, *90*, 215–238.
- McArthur, L. Z., & Ginsberg, E. (1981). Causal attribution to salient stimuli: An investigation of visual fixation mediators. *Personality and Social Psychology Bulletin*, *7*, 547–553.
- McClelland, D. C. (1972). Opinions reflect opinions: So what else is new? *Journal of Consulting and Clinical Psychology*, *38*, 325–326.
- McCrae, R. R. (1982). Consensual validation of personality traits: Evi-

- dence from self-reports and ratings. *Journal of Personality and Social Psychology*, 43, 293-303.
- Milgram, S. (1975). *Obedience to authority*. New York: Harper & Row.
- Miller, A. G., & Rorer, L. G. (1982). Toward an understanding of the fundamental attribution error: Essay diagnosticity in the attitude attribution paradigm. *Journal of Research in Personality*, 16, 41-59.
- Minsky, M. (1975). A framework for representing knowledge. In P. H. Winston (Ed.), *The psychology of computer vision*. New York: McGraw-Hill.
- Mischel, W. (1984). Convergences and challenges in the search for consistency. *American Psychologist*, 39, 351-364.
- Monson, T. C., & Snyder, M. (1977). Actors, observers, and the attribution process: Toward a reconceptualization. *Journal of Experimental Social Psychology*, 13, 89-111.
- Monson, T. C., Tanke, E. D., & Lund, J. (1980). Determinants of social perception in a naturalistic setting. *Journal of Research in Personality*, 14, 104-120.
- Mook, D. G. (1983). In defense of external invalidity. *American Psychologist*, 38, 379-387.
- Moskowitz, D. S., & Schwarz, J. C. (1982). Validity comparison of behavior counts and ratings by knowledgeable informants. *Journal of Personality and Social Psychology*, 42, 518-528.
- Murray, H. A. (1938). *Explorations in personality*. New York: Oxford University Press.
- Nisbett, R. E., & Borgida, E. (1975). Attribution and the psychology of prediction. *Journal of Personality and Social Psychology*, 32, 932-943.
- Nisbett, R. E., Krantz, D. H., Jepson, C., & Kunda, Z. (1983). The use of statistical heuristics in everyday inductive reasoning. *Psychological Review*, 90, 339-363.
- Nisbett, R., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment*. New York: Prentice Hall.
- Norman, W. T. (1969). "To see ourselves as others see us!": Relations among self-perceptions, peer-perceptions, and expected peer-perceptions of personality attributes. *Multivariate Behavior Research*, 4, 417-433.
- Norman, W. T., & Goldberg, L. R. (1966). Raters, ratees, and randomness in personality structure. *Journal of Personality and Social Psychology*, 4, 681-691.
- O'Grady, K. E. (1982). Measures of explained variance: Cautions and limitations. *Psychological Bulletin*, 92, 766-777.
- Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, 17, 776-783.
- Over, R. (1968). Explanations of visual illusions. *Psychological Bulletin*, 70, 545-562.
- Ozer, D. J. (1979, September). *Observer evaluations of personality: 1918-1937*. Paper presented at the 87th Annual Convention of the American Psychological Association, New York.
- Ozer, D. J. (1985). Correlation and the coefficient of determination. *Psychological Bulletin*, 97, 307-315.
- Paunonen, S. V., & Jackson, D. N. (1985). Idiographic measurement strategies for personality and prediction: Some unredeemed promissory notes. *Psychological Review*, 92, 486-511.
- Peterson, C. R., & Beach, L. R. (1967). Man as an intuitive statistician. *Psychological Bulletin*, 84, 231-259.
- Pryor, J. B., Gibbons, F. X., Wicklund, R. A., Fazio, R. H., & Hood, R. (1977). Self-focused attention and self-report validity. *Journal of Personality*, 45, 513-527.
- Quattrone, G. A. (1982). Overattribution and unit formation: When behavior engulfs the person. *Journal of Personality and Social Psychology*, 43, 593-607.
- Rosenthal, R., & Rosnow, R. L. (1984). *Essentials of behavioral research: Methods and data analysis*. New York: McGraw-Hill.
- Rosenthal, R., & Rubin, D. B. (1979). A note on percent variance explained as a measure of the importance of effects. *Journal of Applied Social Psychology*, 9, 395-396.
- Rosenthal, R., & Rubin, D. B. (1982). A simple, general purpose display of magnitude of experimental effect. *Journal of Educational Psychology*, 74, 166-169.
- Ross, L. (1977). The intuitive psychologist and his shortcomings. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 10, pp. 174-214). New York: Academic Press.
- Ross, L., Amabile, T. M., & Steinmetz, J. L. (1977). Social roles, social control, and biases in social-perception processes. *Journal of Personality and Social Psychology*, 35, 485-494.
- Ross, M., & Fletcher, G. J. O. (1985). Attribution and social perception. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology* (3rd ed., Vol. 2, pp. 73-122). New York: Random House.
- Schank, R. C., & Abelson, R. P. (1970). *Scripts, plans, goals, and understanding*. Hillsdale, NJ: Erlbaum.
- Schneider, D. J., Hastorf, A. H., & Ellsworth, P. C. (1979). *Person perception* (2nd ed.). Reading, MA: Addison-Wesley.
- Seligman, M. E. P. (1975). *Helplessness*. San Francisco: Freeman.
- Shafer, G. (1985, February). *Is subjective expected utility normative?* Colloquium delivered to the Department of Statistics, Harvard University.
- Shrauger, J. S., & Schoeneman, T. J. (1979). Symbolic interactionist view of self-concept: Through the looking glass darkly. *Psychological Bulletin*, 86, 549-573.
- Slovic, P., & Lichtenstein, S. (1971). Comparison of Bayesian and regression approaches to the study of information processing in judgment. *Organizational Behavior and Human Performance*, 6, 649-744.
- Swann, W. B., Jr. (1984). Quest for accuracy in person perception: A matter of pragmatics. *Psychological Review*, 91, 457-477.
- Swann, W. B., Jr., Giuliano, T., & Wegner, D. M. (1982). Where leading questions can lead: The power of conjecture in social interaction. *Journal of Personality and Social Psychology*, 42, 1025-1035.
- Taft, R. (1955). The ability to judge people. *Psychological Bulletin*, 52, 1-23.
- Trope, Y., Bassok, M., & Alon, E. (1984). The questions lay interviewers ask. *Journal of Personality*, 52, 90-106.
- Tryon, R. C. (1929). The interpretation of the correlation coefficient. *Psychological Review*, 36, 419-445.
- Tversky, A., & Kahneman, D. (1978). Causal schemata in judgments under uncertainty. In M. Fishbein (Ed.), *Progress in social psychology* (pp. 49-72). Hillsdale, NJ: Erlbaum.
- Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, 90, 293-315.
- Vernon, P. E. (1933). Some characteristics of the good judge of personality. *Journal of Social Psychology*, 4, 42-58.
- Waters, L. K., & Waters, C. W. (1970). Peer nominations as predictors of short-term sale performance. *Journal of Applied Social Psychology*, 54, 42-44.
- West, S. G. (Ed.). (1983). Personality and prediction: An introduction. *Journal of Personality*, 51, 275-285.
- Wiggins, J. S. (1973). *Personality and prediction: Principles of personality assessment*. Reading, MA: Addison-Wesley.
- Winkler, R. L., & Murphy, A. H. (1973). Experiments in the laboratory and the real world. *Organizational Behavior and Human Performance*, 10, 252-270.
- Woodruffe, C. (1985). Consensual validation of personality traits: Additional evidence and individual differences. *Journal of Personality and Social Psychology*, 48, 1240-1252.

Received November 4, 1985

Revision received March 10, 1986 ■