whether one person judges one other person or 1000 people judge 1000 other people.

For example, Archer and Akert (1977) developed a test of nonverbal sensitivity that involved assessing people's ability to accurately identify objective aspects of people, their experiences, or their relationships. In two of their test questions, people were exposed to two women playing with a baby and talking to two men who had just finished a one-on-one basketball game.

Participants' task was to correctly identify the mother of the baby and the winner of the game. In the actual test, there is a series of such vignettes and the more questions answered correctly, the more accurately people are at judging others on this test. This is simple and objective; there is no criterion problem here.

The criterion issue is also a nonproblem in much of the deception-detection literature. That is, one question psychologists have asked is "How good are people at detecting when others lie?" For example, Ekman and Friesen (1969, 1974) had targets view either a pleasant film or a horrible gory one. All had to inform perceivers that they had just seen a lovely, pleasant film. The research question was how often people can tell when others are liars. Again, there is no criterion problem.

In studies of some aspects of stereotypes, one can also compare people's perceptions to objective data. In one classic study, McGuire and Stitt (1978) compared people's beliefs about differences between African Americans and other Americans to U.S. Census data (e.g., in size of family and likelihood of completing high school). In others, people's beliefs about sex differences (e.g., in aggression) were compared to the results from meta-analyses assessing sex differences (Hall & Carter, 1999; Swim, 1994). Although even the U.S. Census and meta-analyses do not necessarily reveal Absolute Truth, they do provide strong sources of evidence.

With this type of data, one can establish overall levels of accuracy, error, and bias. Do people err on the side of believing what others say? Then Ekman and Friesen's studies should have yielded results showing that people underestimate deception (it did, but their research also typically showed that people do better than chance at accurately detecting deception).

Do people exaggerate differences between demographic groups? Then McGuire and Stitt's (1978) ans; Swim (1994) should have found people consistently overestimating the differences between different groups (they did not; people tended to be accurate and, when wrong, to underestimate real differences). These are important and interesting accuracy questions, and the availability of clear, objective criteria in these cases is not problematic.

3. Independent and Standardized but Not Universally Persuasive Objective Criteria

Not all people may agree that certain objective criteria are good ones. Such agreement might be irrelevant regarding a target's height or marital status, but it becomes much more relevant when estimating a target's extra-

version or intelligence via a personality questionnaire or standardized IQ test. Is the personality questionnaire a good one? Is it reliable? Valid? IQ tests, in particular, have a long and controversial history (e.g., Gould, 1973; Herrnstein & Murray, 1994; Neisser, Boodoo, Bouchard, Boykin, Brody, Ceci, Halpern, Lohchin, Perloff, Sternberg, & Urbina, 1996), as has the measurement of enduring personality traits (see Funder, 2001, for a review).

To the extent that some people do not find such tests credible, they are likely to discredit or dismiss research on accuracy using such criteria. Thus, use of objective but controversial criteria can be viewed as holding down agreement involving criteria the study assesses accuracy; if one does not agree with the criteria, the study assesses accuracy; if one does not agree with the criteria, the study assesses accuracy; if one does not agree with the criteria, the study assesses accuracy.

Often, however, the reverse may happen: People who do not like a study's conclusions will come up with arguments against the appropriateness of its criteria (Lord, Roof, & Lepper, 1979). Indeed, people are often extremely good at engaging in the intellectual contortions necessary to maintain a cherished viewpoint even in the face of clear evidence that they are wrong (Aronson, 1999; Festinger, Riecken, & Schachter, 1956). The fact that some people do not like or respect some criterion, of course, does not mean that it cannot be used. It only means that one needs to make a credible case for the appropriateness of that criterion in one's study.

For example, standardized test scores have often been used as criteria in studies of the extent to which teacher expectations for students were self-

fulfilling prophecies or accurate (see edited volumes by Brophy, 1996; Duckel, 1985; see reviews by Brophy, 1983; Brophy & Good, 1974; Jusin, 1986; Spitz, 1999). If teachers are good at identifying which students are smarter, then their expectations should predict but not influence students' standardized test scores. If teacher expectations do predict standardized test scores without causing them, any real or imagined weaknesses in standardized tests would seem largely irrelevant to interpreting this result.

Use of imperfect criteria will generally lead researchers to underestimate people's accuracy. If so, in sharp contrast to attempts to dismiss the viability of accuracy research because of imperfect criteria, this means that people are

NOTICE
This material may be protected by Copyright Law (Title 17 U.S. code)
probably even more accurate than indicated by the evidence emerging from any particular study using imperfect criteria. This is obviously the case when using a criterion measure with less than perfect reliability. Unreliability in measurement artificially lowers correlations between judgments and criteria, so that the greater the unreliability, the higher the actual accuracy. Consider a case in which people’s perceptions of others’ extraversion correlate .3 with a self-report extraversion scale. If there is less-than-perfect reliability in measurement of both perception and criterion, then the true correlation will be higher than .3. If, for example, the reliability of both scales was .9, the true correlation between perception and criterion would be .33; if the reliability of both scales was .75, the true correlation between perception and criterion would be .4 (Carmines & Zeller, 1979).

This would often be equally true with a scale of imperfect or partial validity. Perhaps a particular psychological measure only captures a partial aspect of some attribute. People’s judgments, perceptions, and so forth, if they reflect more aspects of the attribute than the measure being used, may be more accurate than indicated by the correlation between judgment and criterion. Consider an assessment of people’s accuracy in judging how big a target is. The criterion, however, is height. Height, of course, is only one aspect of size (the others being width and depth). If people use height, width, and depth to judge size, the correlation of their judgments with the (imperfect, partial) criterion will be too low. That means their accuracy could be even higher than indicated by that criterion.

Similarly, consider the use of an IQ test as criterion for evaluating the accuracy of judgments of intelligence. If a particular IQ test primarily taps verbal intelligence and people’s judgments include, in addition to estimations of verbal intelligence, creativity, social skill, political savvy, wit, and common sense, people’s judgments closely correspond to actual intelligence defined in this multifaceted way, the correlation between judgments and criterion will be artificially low. People might be more accurate than indicated by the correlation (correspondence) between judgment and criterion. An identical analysis applies to using any measure, such as a standard personality inventory, as a criterion for assessing accuracy in judgments of personality traits.

Of course, just because such mismatches may lower the observed correlation between perception and criterion, one cannot infer high accuracy from low correlations, but the argument that any particular criterion in any particular study is “bad” (unreliable, low validity) means that whatever evidence on accuracy is obtained probably often represents a lower bound on accuracy. This, in turn, means that people may actually be more accurate than indicated by the empirical evidence of accuracy obtained in that study. Just as imperfect criteria do not preclude the possibility of 

ACCURACY IN SOCIAL PERCEPTION

objective, social scientific research on all sorts of topics (aggression, identity, achievement, social class, etc.), nor do they preclude the possibility of accuracy research. Some criteria are subject to debate and controversy. The extent to which nearly all measures of social science constructs have been validated or well established has limitations, thereby inflicting imperfections on any research using any measure of any construct. Researchers should almost always carefully evaluate the flaws and limitations of their research, including their use of measures. However, this issue is no more a problem for accuracy research than for any other type of research.

E. BEHAVIOR

1. What Is Behavior?

In the broadest sense, anything a person (or any organism) does is behavior. In this sense, answers on a standardized test constitute behavior. Similarly, agreement among judges usually constitutes agreement concerning other targets’ behavior or inferences about their attributes based on their behavior (Kragulski, 1989). Even completely unobservable thinking, feeling, or daydreaming can be considered behavior. Indeed, transmission or inhibition of neurotransmitters can be considered behavior at the level of the individual cell.

None of this is what is meant by behavior here. For the social perception context addressed in this chapter, behavior is considered to be observable action, not unobservable thought processes, cell electrochemical processes, or underlying attributes. Counting how frequently Natasha smiled is related to but different from determining whether she is thinking happy thoughts or how sociable she is. Smiles are visible. Similarly, measuring how much Jane Jones spends on some task is related to but different from evaluating Jane’s motivation. Again, time on task is an observable behavior, not an underlying attribute.

Similarly, “answers on a standardized test” is not what most social scientists mean when they use the term “behavior.” Such tests are generally designed to measure some attribute that has relevance beyond the testing situation. Intelligence test scores should predict school outcomes, work outcomes, and many life outcomes. Scores on a personality scale measuring hostility should predict levels of angry and aggressive behavior in all sorts of interpersonal contexts. Although standardized test scores could be considered one class of behavior, it is both possible and useful to distinguish between standardized tests and overt actions.
2. Does Behavior Bail Down to Agreement?

Some have argued that it does (e.g., Kuglinski, 1989), and for good reason. At least in most research contexts, a common way to measure human behavior is by having independent judges observe and code the behavior. However, there is a difference between the need for people to observe and record behavior and the need for people to interpret behavior. Behavior itself is observable; personality, beliefs, attitudes, motivation, competence, etc., can be inferred from behavior, but they are not directly observable. Observers recording nonverbal behaviors (e.g., smiling, speech dysfluencies, etc.), time spent on some task (e.g., proportion of time spent talking in a dyadic conversation; time spent trying to solve an impossible anagram before giving up; time it takes to walk down a hallway, etc.), task choice (how many advanced math courses does a person take? Do they apply to college? Number of parties attended, etc.) are all recording objective aspects of behavior.

Around the fringes, there may be room for interpretation. Was Fred smiling, or was the corner of his mouth just twitching a bit? And it is very useful to have multiple observers record behavior in order to resolve these fringe or ambiguous situations. When that happens, judgments of behavior may be down to agreement. Nonetheless, when behavior rather than underlying trait is being recorded, such fringe or ambiguous situations are generally likely to be the exception rather than the rule. Thus, there is an important distinction between using observers to record what a person does and using observers to interpret and evaluate the meaning of what a person does. Behavior is what a person does.

Many of the classic programs of research in social psychology have studied phenomena with behavioral criteria, such as research on the attitude-behavior relationship (Eagly & Chaiken, 1993), self-efficacy (Bandura, 1977), behavioral confirmation (Snyder, 1984), obedience and conformity (Asch, 1955; Milgram, 1974), and helping in an emergency (e.g., Latane & Darley, 1970). I know of no critical work that takes these areas to task for their use of behavioral criteria as somehow "problematic" or even for "merely studying agreement." The rationale for believing that behavior can be identified when one studies such classic social psychology topics as these, but not when one studies accuracy, has never been explicitly stated in any prior article criticizing or rejecting accuracy research or arguing that accuracy boils down to nothing more than agreement (e.g., Claire & Fish, 1998; Fish, 1998; Jones, 1986, 1990; Kuglinski, 1989; Stagner, 1995). It remains a challenge, therefore, for scientists arguing that it is somehow problematic to use behavior as a criterion for studying accuracy to explain why the use of behavioral criteria is not equally problematic in other areas of psychology.

ACCURACY IN SOCIAL PERCEPTION

F. AGREEMENT WITH OTHER PERCEIVERS

Agreement, at first glance, appears to be a poor criterion for accuracy. Agreeing that the world is flat does not make it flat. Thus, agreement, like other criteria, is almost always imperfect.

But agreement cannot be so readily dismissed. If both of us are accurate, we must also agree. If a baseball player hits a home run, and we both saw it, we will both agree that he hit a home run. If Banna is a brilliant student, and we both recognize her brilliance, we will both agree, intelligence, and so forth may be inferred from behavior, but they are not directly observable. Observers recording nonverbal behaviors (e.g., smiling, speech dysfluencies, etc.), time spent on some task (e.g., proportion of time spent talking in a dyadic conversation; time spent trying to solve an impossible anagram before giving up; time it takes to walk down a hallway, etc.), task choice (how many advanced math courses does a person take? Do they apply to college? Number of parties attended, etc.) are all recording objective aspects of behavior.

Around the fringes, there may be room for interpretation. Was Fred smiling, or was the corner of his mouth just twitching a bit? And it is very useful to have multiple observers record behavior in order to resolve these fringe or ambiguous situations. When that happens, judgments of behavior may be down to agreement. Nonetheless, when behavior rather than underlying trait is being recorded, such fringe or ambiguous situations are generally likely to be the exception rather than the rule. Thus, there is an important distinction between using observers to record what a person does and using observers to interpret and evaluate the meaning of what a person does. Behavior is what a person does.

Many of the classic programs of research in social psychology have studied phenomena with behavioral criteria, such as research on the attitude-behavior relationship (Eagly & Chaiken, 1993), self-efficacy (Bandura, 1977), behavioral confirmation (Snyder, 1984), obedience and conformity (Asch, 1955; Milgram, 1974), and helping in an emergency (e.g., Latane & Darley, 1970). I know of no critical work that takes these areas to task for their use of behavioral criteria as somehow "problematic" or even for "merely studying agreement." The rationale for believing that behavior can be identified when one studies such classic social psychology topics as these, but not when one studies accuracy, has never been explicitly stated in any prior article criticizing or rejecting accuracy research or arguing that accuracy boils down to nothing more than agreement (e.g., Claire & Fish, 1998; Fish, 1998; Jones, 1986, 1990; Kuglinski, 1989; Stagner, 1995). It remains a challenge, therefore, for scientists arguing that it is somehow problematic to use behavior as a criterion for studying accuracy to explain why the use of behavioral criteria is not equally problematic in other areas of psychology.

We do want to be careful in considering what constitutes "the same thing." Yes, and I could disagree about how pleasant, extraneous, or intelligible Fred is. If Fred acts differently with you from with me, we both could be right (Sears, 1984). So if we are talking about Fred in general, both of us may be partially right and partially wrong, because perhaps there is no "Fred in general" — there is only Fred interacting differently with different people.
might be compared with those of the students' teachers. For some types of research, nonprofessionals could also be considered "experts" in the sense that they have some sort of unique access to knowledge about the target. Thus, spouses, romantically involved couples, close friends, and coworkers could be used as experts with whom to compare perceivers' judgments of targets.

2. Agreement with Experts' Models

Sometimes, experts, such as statisticians, decision-making theorists, or psychologists, have developed formal models for what constitutes the most appropriate way to arrive at a judgment. Accounting for regression to the mean, appropriate use of base rates, and use of common sense, consistency, and distinctiveness in arriving at attributions all constitute formal models of the appropriateness of social judgments (see, e.g., Dawes, 1979; Kahneinan et al., 1982, Kelley, 1967; Nisbett & Ross, 1980). Thus, judgments or decisions based on such models may be used as criteria against which to evaluate the accuracy of laypeople's judgments or decisions.

3. Agreement with Independent Judges

If a social interaction is recorded, for example, by videotape, transcription, and so on, or if it can be monitored (e.g., through intercoms, computers, or one-way mirrors), people other than the perceiver can be asked to evaluate the target. These people are referred to here as "independent" judges because they were not involved in the interaction between perceiver and target; thus, they are independent of the perceiver. Often in such situations, these independent judges may only be exposed to the targets' responses to further maximize any effects of the perceiver's verbal or nonverbal behavior on the independent judges (e.g., Goldman & Lewis, 1977; Word et al., 1974). Thus, the perceptions of independent judges may be used as criteria against which to evaluate the accuracy of the perceiver's person perception judgments.

4. Agreement with Nonindependent Judges

Sometimes the judgments of people who do interact with the target, either simultaneously while interacting with the perceiver or in other contexts, may be used as criteria for evaluating the perceiver's accuracy. In such cases, all perceivers' judgments of targets essentially constitute the criteria for one another. Indeed, Kenny's (1964) social relations model, which acknowledges numerous aspects of agreement, accuracy, and social perception, requires multiple people to interact with and rate one another. Nonindependent judges may be strangers, acquaintances, or experts.

5. Limitations to Agreement with Other Perceivers

The perspective of probabilistic realism indicates that nearly all criteria will be imperfect to some degree. Agreement is no exception. The fundamental limitation to use of agreement is, of course, that everyone can be wrong.

a. Even Experts May Be Wrong. Indeed, sometimes, expert judgments or predictions may be little better than those of any people (Cronbach, 1955). Psychiatrists, for example, predicted that hardly any teachers would give 450 volts worth of shock to learners providing wrong answers on a test in Milgram's (1974) studies of obedience (even though half or more often did). In addition, experts may be subject to their own biases. Happy romantic couples, for example, may see each other in an overly positive, almost idealized way (Murray, Holmes, & Griffin, 1996). The doctors and staff at psychiatric hospitals at least sometimes misinterpret patient boredom or hostility as psychopathology (Rosenthal, 1973). Teachers' expectations sometimes color their interpretations of students' performance (Williams, 1976). Thus, expert judgments are susceptible to a variety of potential imperfections. They are probably best used as a criterion for accuracy when prior research has provided strong evidence for their validity.

b. Experts' Models May Be Contradictory. Expert model often have a formal logical rigor or statistical/mathematical basis that gives them immediate credibility. Nonetheless, they may not always be correct. This will be demonstrated below using two examples in which different expert models make opposing predictions (meaning that both cannot possibly be right):

- one involving the stock market and the other involving stereotypes.

Efficient markets theory (the dominant view of the stock market among economists) states, in essence, that because whatever is known is already factored into stock prices, it is impossible to consistently obtain returns that beat the overall market (Malkiel, 1999). The stock market, according to this view, is essentially a random walk (the future is inherently unpredictable) with an upward overall trajectory. The practical implication is that most people who do not plan to spend their investment dollars any time soon will receive the greatest return by buying a stock index fund that reflects the entire stock market and holding through (unpredictable) ups and downs. One will most likely get oneself into trouble (reduce one's overall return) if one tries to "time" the market (buy low, sell high) or select individual stocks to beat the market averages.
Efficient markets theory, however, can be viewed as clashing with one of the main claims of statistical experts. The statistical idea of regression to the average means that extreme values are likely to return over time to the overall mean. This could be interpreted to mean that if stock valuations are unusually high, one is more likely to lose money and less likely to make money (going forward) than usual; and if they are unusually low, one is more likely to make money and less likely to lose money (going forward) than usual. If one enters the market when valuations are unusually low, therefore, and exits when they are unusually high, one should be able to beat the overall market (by receiving the unusually large gains that follow low valuations and avoiding the severe losses following high valuations).

ii. Bayesian inference versus stereotypes "biasing" person perception.

A similar contradiction occurs in expert models regarding use of stereotypes. Many social psychological theories and perspectives on stereotypes, much current cultural discourse, and many laws and judicial decisions state or implicitly assume that people are acting reasonably and appropriately only when they judge others solely and entirely on the basis of those others' personal characteristics, rather than their group memberships. This view argues or implies that people act in a biased, prejudiced, or irrational manner when they allow their stereotypes to influence their judgments of individuals (e.g., Borgida et al., 1993; Brewer, 1988a; Darley & Fazio, 1980; Dunning & Sherman, 1997; Fiske & Neuberg, 1990; Jones, 1986, 1990; Myerson, 1999; Nelson, 2002).

According to widely accepted principles of probability (e.g., Bayes's theorem, see e.g., Kahneman & Tversky, 1973, regression (e.g., Jussim, 1991), and philosophy of science (e.g., Krueger & Funder, 2004; Meehl, 1990), however, base rates, prior beliefs, and expectations should often influence people's interpretations of new evidence (see McCauley et al., 1980 for an analysis focusing specifically on stereotypes). These principles suggest that if perceivers are not sure how tall a person is, they will, on average, be right more often if they estimate any particular male to be a few inches taller than any particular female than if they estimate them to be exactly equal (i.e., use, rather than discard, their sex stereotype regarding height). Even if perceivers find out that both a male and female target are avid basketball players (identical individuating information), the best guess, according to these principles, is still that the male is taller than the female. Only if the perceivers find out that both are 6 feet tall, which is perfectly diagnostic of height, should they not apply their belief that men are usually taller than women.

This set of alternative expert models has created the following amusing and ironic state of affairs within the stereotypes literature. Specifically, people are accused of irrationality if they ignore their stereotypes when judging individual targets because they are supposed to use base rates when judging other people under uncertainty (e.g., Kahneman & Tversky, 1973; Locksley, Borgida, Brekke, & Hehrens, 1980); and more typically, they are accused of error and bias if they use their stereotypes when judging individual targets (e.g., American Psychological Association, 1991; Brewer, 1988a; Darley & Gross, 1983; Dunning & Sherman, 1997; Fiske, 1998; Fiske & Neuberg, 1990; Jones, 1986, 1990; Nelson, 2002). These and other discussions of the role of stereotypes in person perception assume that people are acting in an irrational, biased, or prejudiced manner if they perceive mass differences (on whatever attribute or behavior is being judged) between individuals belonging to different social groups holding individuating information constant (Bodenhorn et al., 1988; Borgida et al., 1993; Darley & Fazio, 1980; Hamilton et al., 1990; Miller & Turnbull, 1986). Relying exclusively on individuating information and ignoring stereotypes base rates is often held up almost as a moral ideal to which people should strive.

It cannot be normatively appropriate to both use and ignore prior expectations. Thus, the expert assumption that people should judge others solely on the basis of individuating information, which has emerged from much of the stereotype literature, conflicts with the claim that prior expectations and beliefs should influence judgment (in the absence of perfectly diagnostic individuating information) that have emerged from the cognitive judgment and decision-making literature.

In general, explicitly stating one's model of rationality is typical of the judgment and decision-making literature (e.g., Kahneman et al.), 1982) but atypical of the stereotyping literature. The few exceptions in the stereotype literature (e.g., Krueger, Hasman, & Auvree, 2005; Locksley et al., 1990; McCauley et al., 1980) use Bayes's theorem as their model of rationality. Obviously, by adopting the same model that is used to explain much of the judgment and decision-making literature, these perspectives avoid the conflict between expert models concerning stereotypes.

However, apparently, there are social and political risks in scientists relying on Bayes's theorem when attempting to understand accuracy, error, and bias in judgments of individuals. When McCauley et al. (1995) argued that a Bayesian analysis indicates that reliance on accurate beliefs about groups should increase, not decrease, the accuracy of person perception judgments in the absence of perfectly diagnostic individuating information, they were accused of "disregarding with civil rights law" (Onsley, 1998, p. 381). Such an accusation conflates moral and legal issues with accuracy; if valid, represents another situation in which it may be illegal to be accurate, and highlights the role of politics in some critiques of stereotype accuracy research.

Regardless, discussions emphasizing the power of stereotypes to bias judgments (e.g., Aronson, 1999; Borgida et al., 1993; Darley & Fazio, 1980;
G. AGREEMENT WITH THE TARGET

Agreement with targets' self-descriptions can and have been used as a criterion for assessing accuracy. Two broad types of targets' self-descriptions are distinguished here. The term "self-reports of behavior" is used to refer to the actions targets claim they have engaged in. Examples might be how many glasses of alcohol they consumed yesterday, how much time they spend exercising each week, how often they argue with their spouse, and how much sleep they get each night. In contrast, the term "self-perceptions" is used to include targets' attitudes, beliefs, feelings, and evaluations of themselves, their characteristics, and their accomplishments. Self-perceptions might include self-esteem, self-perceptions regarding personality traits (independent, assertive, extraverted, etc.), feelings, political positions, self-evaluations of academic or athletic performance, and so on. Self-perceptions, therefore, generally involve unobservable, underlying attributes of some type, whereas self-reports of behavior involve overt, observable actions.

1. Agreement with Targets' Self-Reports of Behavior

Targets' self-reports of behavior can be used as a criterion against which to evaluate perceivers' judgments. Social reality typically constrains bias (e.g., Junissim, 1991; Kunda, 1990) so that self-reports regarding specific and objective behaviors may be less likely to be tainted by self-serving.

ACCURACY IN SOCIAL PERCEPTION

defensive, or impression management biases than are self-reports regarding vague or ambiguous attributes.

To get concrete, a fraternity member's response to "How many alcoholic drinks have you had in the last week?" may be less likely to be biased than is his response to questions such as "How often did you get drunk last week?" Although bias may emerge regarding even the most objective of behaviors, there is a lot more room for interpretation in the "drunk" question ("well, I did drink two sixes of beer, but I never really got drunk") than in the "drinks" question. Similarly, responses to "How much money did you donate to charity last month?" are more likely to be constrained by reality than are responses to "How generous are you?" Again, bias is always possible, but even a highly biased person is not likely to interpret buying a pizza as donating to charity, although people may vary a great deal on whether they consider sharing the pizza with a friend as a hallmark of great generosity.

2. Agreement with Targets' Self-Perceptions

Despite their imperfections, targets' own self-perceptions regarding underlying or ambiguous attributes, such as personality characteristics or dispositions, can often be used as a criterion (see e.g., Judd et al., 1995; Ryan, 1995). As usual with probabilistic realism, the issue is not whether such a criterion is perfect, because no criteria are perfect. The question is whether the specific self-perceptions being used as a criterion are likely to reflect what the target is like.

Indeed, there are theoretical reasons for expecting self-perceptions to be good criteria, at least sometimes. People have access to much more information about some of their experiences, inner states, relationships, and so on than do outsiders. Indeed, at least some research has shown that people's memory for both randomly sampled and particularly striking thoughts and actions, up to 5 months after they occurred, is excellent (e.g., Bacles & DeCooke, 1988; Brewer, 1985).

Because bias can exist aside by side with accuracy (e.g., Junissim, 1991), even when biases taint self-perceptions, they do not necessarily eliminate their validity (see, e.g., Hend, 1965; Funder, 1993). If so, they can be used in the same manner as any imperfect criteria. Self-perceptions of academic ability, for example, usually correspond highly with indicators of academic achievement, such as grades and standardized test scores (e.g., Eccles & Wigfield, 1985; Selmon, 1984). Attributes that are more readily observable, such as sociality or extraversion, tend to generate high agreement between self-reports and observers' ratings (e.g., Funder, 1995; Kenny, 1994).
Furthermore, considerable validity evidence has been obtained regarding numerous self-report scales of personality characteristics and political attitudes (Robinson, Shaver, & Wrightsman, 1991, 1999). More highly validated self-perceptions are, obviously, better criteria than less well validated self-perceptions. If well-validated by prior research, however, self-report scales cannot (or at least should not) be summarily dismissed as ‘miscellaneous’ self-report, at least not without inflating the preexisting validity evidence for the scale. In the absence of clear validity evidence, however, self-reports must be interpreted with caution and due regard to the topic. Self-reports of height, broccoli-eating, and interest in pornography fall on a continuum of least to most likely to be contaminated by social desirability biases.

3. Limitations to Agreement with Targets' Self-Reports and Self-Perceptions

Targets, of course, are imperfect themselves. Although they may have unique access to certain types of information (personal experiences, feelings, etc.), many people are subject to both motivated and unmotivated errors and biases (Kunda, 1990; Niët & Ross, 1980). Sometimes, people do not have access to their own cognitive processes or to some of their implicit beliefs and attitudes (Nisbett & Wilson, 1977; Wilson, Lindsey, & Schooler, 2000). Memory is imperfect and potentially subject to all sorts of biases (e.g., Kotar et al., 2000; Stanger & McMillan, 1992), so that self-reports may often be imperfect records of behavior. Researchers, however, have also developed a slew of methods for improving the accuracy of self-reports (e.g., event sampling, daily diary methods, etc.). Although even these types of methods may not completely eliminate error and bias, they have become widely used because of their demonstrated validity for many types of self-reports (e.g., Hedge, Jandorf, & Stone, 1985; Räikkönen, Matthews, Flory, Owens, & Gump, 1999).

The problem of motivated biases may often (though not always) be greater when using self-perceptions (rather than self-reports of behavior) as a criterion, in part because of the common tendency for most people to view themselves in socially desirable, moral, and competent a manner as possible (Paulhus, 1991, 2002). Although this means that the overall average level for some self-perception may often be too favorable, this may not prevent self-perceptions from being a good correlational criterion for assessing the accuracy of perceivers' beliefs.

ACCURACY IN SOCIAL PERCEPTION

For example, Bertha may think she is a great athlete and Nyesha may think she is a good athlete. Both may be overestimates (Bertha may only be pretty good and Nyesha may be pretty average), but if their degree of self-inflation is similar, it may be true that Bertha is more athletic than Nyesha. So a coach who views Bertha as more athletic than Nyesha would be correct (and the coach's views would correlate well with Bertha and Nyesha's self-perceptions). Thus, even biased target self-perceptions may, under many circumstances, constitute good, if imperfect, criteria for assessing the accuracy of perceivers. In practice, although a default rejection of self-perceptions as criteria is not justified, researchers considering the use of particular self-perceptions need to thoughtfully consider their limitations and develop a convincing case that, in the context under study, they are likely to be good criteria.

H. CRITERIA: CONCLUSIONS

Probabilistic realism occasionally provides a gold standard for establishing accuracy. A real estate salesman either did or did not sell the house. Except in rare cases, that is not a matter of opinion. The absence of such a gold standard, however, does not justify the conclusion that the criteria available to assess accuracy are so fundamentally flawed or their interpretation so ultimately ambiguous as to cloud the meaning of accuracy research. Establishing that a social belief or perception is accurate is much like establishing validity in social science research.

The strongest and clearest evidence regarding accuracy comes from research that typically uses multiple measures and methods to establish the accuracy of social perception (e.g., Cronbach, 1955; Funder, 1987, 1995; Kenny, 1994). This does not mean that research examining accuracy using a single method or criterion is uninformative. Such research, however, may be less informative than research using multiple methods or criteria (except when that single criterion itself has previously been validated using multiple criteria).

Despite scientists' inability to achieve absolute 100% certainty, probabilistic realism, truth with a small "t," and the principles of construct validity provide clear ways of identifying criteria against which researchers can assess the accuracy of social perception. Once one has a good set of criteria, assessing accuracy would appear to be simple—just evaluate how well the perceptions correspond with the criteria. In fact, however, this turns out to be considerably more complex than it seems at first glance, and the next section explains why.
VIII. Must All Research on Accuracy Assess Components?

A stopped clock is right twice each day. That does not make it a good clock. What does this have to do with social perception? More than it seems. Consider the following example: A wife successfully predicts that her husband will come home at least half an hour later than he said he would come home. It looks like this wife knows her husband pretty well, doesn’t it? Not necessarily. Maybe she always predicts that her husband will be late. Maybe she always predicts everyone will be late. It could even be worse than that. Maybe she always predicts all people will always do bad things and have bad attributes.

Even though she might have happened to be right that one time, she could not necessarily be considered a particularly astute judge of her husband. One could think of her prediction as stemming from several sources or components: her overall tendency to think well (or poorly) of people, her overall tendency to predict that people will be late (over and above her general tendency to think well/poorly of people), her overall tendency to think that her husband will be late (over and above her general tendency to think well of people and that people will be late), and her specific tendency to predict that her husband will come home later than he claimed this particular time (over and above her tendency to think well of people; to predict that they, in general, will be late; and to think her husband will run late). Each component of her prediction can be accurate to some degree, and each contributes to her overall likelihood of being accurate (across lots of judgments or predictions).

A. COMPONENTIAL APPROACHES TO THE STUDY OF ACCURACY

This type of thinking inspired Cronbach’s (1955) classic review and at least two other more recent perspectives (Judd & Park, 1993; Kenny, 1994), all of which identified several processes contributing to social perception and that argued that accuracy needs to be separated into different components reflecting these different processes. Next, therefore, a brief overview of each of these three componential approaches to the study of accuracy is presented.

1. Cronbach’s Components

Cronbach’s (1955) analysis involves computing an accuracy score for each perceiver judging a set of targets on a set of traits. Thus, there is a separate score for each perceiver. In this situation, both social judgment and social accuracy can be divided according to effects for targets and traits (Kenny, 1994).

Social Judgment = Constant (detection) + Target main effect (differential detection) + Trait main effect (stereotype accuracy) + Target x Trait interaction (differential accuracy).

The constant (which Cronbach called “detection”) is simply the grand mean of all judgments, the target main effect (which Cronbach called “differential detection”) refers to mean differences in how each target is judged (over all traits), the trait main effect (which Cronbach called “stereotype accuracy”) refers to mean differences in how each trait is judged (over all targets), and the target x trait interaction (which Cronbach called “differential accuracy”) refers to how each target is rated on each trait after subtracting out the other three effects.

Social judgments are then compared to criteria, which are broken down into the same sets of components. Accuracy, then, refers to how well each component of the judgment corresponds to the identical component on the criterion. If, for example, a student tends to see other people through rose-colored glasses, her elevation score will be higher than the elevation score on the criterion. If she sees Professor Smith as a better teacher than Professor Jones and the criteria also indicate that Professor Smith is a better teacher than Professor Jones, her differential elevation accuracy will be high even if seeing the world through her rose-colored glasses leads her to overestimate both of them.

If, when rating a group of professors, she sees absent-mindedness as more common than is organization, when in fact few professors are absent-minded but many are well-organized, her stereotype accuracy will be low. And if she rates Jones as more organized than Smith (after subtracting...
out elevation, differential elevation, and stereotype accuracy), and Jones really is more organized than Smith, her differential accuracy will be high (see Kenny (1994, pp. 117-121) for a fully explicated mathematical example of Cronbach's components involving three perceivers, three targets, and three traits).

2. Kenny's Social Relations Model

The Social Relations Model (SRM) is a compositional model that is related to, but differs from, Cronbach's (1955) in several important ways (see, e.g., Kenny, 1994; Kenary & Albright, 1987, for detailed discussions of similarities and differences). First, it is intended to be a broad and general model for assessing many different aspects of social perception; of which accuracy is only one. Second, research using SRM typically focuses on perceptions regarding one trait at a time, rather than the multiplicity of traits addressed by Cronbach's components. Third, however, it also typically focuses on several perceivers, rather than the one perceiver (at a time) that was the focus of Cronbach's analysis. Thus, SRM research might perform one analysis to find out how accurately Dave, Charls, and Bella perceive each other's intelligence and another to find out how accurately they perceive each other's friendliness. Whereas Cronbach partitioned he judgment into target, trait, and target x trait components, SRM partitions judgment into target, perceiver, and target x perceiver components. Thus, the SRM equation is:

Each Social Judgment = Constant (elevation) + Perceiver main effect (perceiver effect) + Target main effect (target effect) + Perceiver x Target interaction (relationship effect). (2)

The equation can be computed separately for each judgment provided by each perceiver, as shown by this example (from Kenny, 1994, p. 130):

Al's judgment of Bob = Constant + Al's perceiver effect + Bob's target effect + Al's relationship with Bob. (2a)

The constant in the grand mean of all judgments, the perceiver main effect is the mean difference between perceivers (over all targets), the target main effect is the mean difference between targets (over all perceivers), and the perceiver x target interaction constitutes the unique manner in which a particular perceiver judges a particular target after subtracting out the constant, the perceiver main effect, and the target, main effect.

In contrast to Cronbach's (1955) components, however, SRM has a substantially different equation for the criterion:

Criteron/Social Reality/Behavior = Constant + Partner effect + Actor effect + Partner x Actor Interaction (3)

This equation can be computed separately for each actor (also from Kenny, 1994, p. 130):

Bob's behavior with Al = Constant + Bob's Actor Effect + Al's Partner effect + Bob's relationship effect with Al. (3a)

The constant is the mean level of the criterion; the actor effect refers to mean differences between actors (averaging over all partners); the partner effect refers to mean differences in how all actors, on average, behave with each partner; and the partner x actor interaction refers to the unique manner in which a particular actor behaves with a particular partner after subtracting out the constant, partner effect, and actor effect.

To assess accuracy, each component of the judgment is compared to the corresponding component on the criterion. Overall, perceivers may see a high level of intelligence in their partners if, on average, the actors really are pretty smart; elevation accuracy will be high (the constant will be similar). If, on average, Al sees his partners as much smarter than they really are, Al's perceiver accuracy will not be high (comparison of Al's perceiver effect [Eq. 2a] to Al's partner effect [Eq. 3a]). If, on average, people see Bob as smarter than the others, and he really is, Bob's target accuracy (called "generalized accuracy" in SRM) will be high (comparison of Bob's target effect, i.e., how he is usually perceived [Eq. 2a], with Bob's actor effect, i.e., what he is usually like [Eq. 3a]). If, on average, Al sees Bob as particularly smart (after subtracting out the constant, Al's perceiver effect, and Bob's target effect from the judgment) and Bob really does act particularly smart when interacting with Al (e.g., as might occur if Al's perception of Bob is self-fulfilling), then Al's dyadic accuracy will be high (comparison of the relationship effects in Eqs. 2a and 3a).

*Although I do not consider self-fulfilling beliefs to be accurate, they could be considered a form of accuracy within SRM.
3. Judd and Park's Full Accuracy Design For Research
On Stereotypes

Judd and Park (1993) developed the first model focused on explaining sources of accuracy and inaccuracy in social stereotypes. Their approach was primarily designed to analyze accuracy and bias in at least two groups' perceptions of one another's attributes. Thus, their analysis incorporates four main components of judgments regarding groups: elevation, perceiver group, target group, and attributes and all possible two-way and three-way combinations of these components.

\[
\text{Judgment} = \text{constant} \times \text{perceiver group effect(pee)} + \\
\text{target group effect(tge)} + \\
\text{attribute effect(ate)} = \\
\text{(pee \times tge)} + \text{(tge \times ate)} + \text{(pee \times ate)} + \\
\text{(pee \times tge \times ate)}.
\]

(4)

Because Judd and Park's (1993) component model is not identical to that of Cronbach's and Kenny's (1994) approach, only a brief simplification of their ideas is presented. The constant is, once again, the mean level of judgments. The perceiver group effect refers to mean differences in the judgments of different groups of perceivers, the target group effect refers to the mean difference in how different groups of targets are judged, and the attribute effect refers to mean differences between attribute classes. The various two- and three-way interactions represent unique combinations of perceiver group, target group, and attribute, over and above the constant and the three main effects. These components are then compared to groups' scores on the criterion. If one group of perceivers generally tends to overestimate targets' score on the criterion, there will be some perceiver elevation inaccuracy. If one group of targets is consistently overestimated, there will be some target elevation inaccuracy. If one class of attributes is consistently overestimated, there will be an attribute elevation inaccuracy. And so on.

Do people see other groups in general in a stereotypical manner? The attribute elevation effect represents an overall tendency to over- or underestimate the prevalence of a particular type or class of attributes across the target groups. If attributes are chosen such that they are stereotypic for one group and counterstereotypic for the other, the attribute elevation effect becomes a "stereotypically" effect—the tendency to view groups as more or less stereotypic than they appear on the criterion. A general tendency to overestimate stereotypical attributes and underestimate counterstereotypical attributes represents a general tendency (across target groups) for the stereotype to exaggerate real differences. A general tendency to underestimate stereotypical attributes and underestimate counterstereotypical attributes represents a general tendency (across target groups) for the stereotype to underestimate real differences.

For example, let's say perceivers are asked to rate men and women on the traits aggressive, strong, and mathematical. High attribute elevation means that people perceive a greater difference between men and women, across these traits, than really exists. This approach can also be used to assess value inaccuarees (is a group seen more positively or negatively than they deserve?) and dispersion inaccuarees (is a group seen as more or less variable than they really are?).

Like the other componential approaches, Judd and Park's (1993) full accuracy design was modeled after an analysis of variance—and with three ANOVA factors (perceiver group, target group, attributes and all two-way and three-way interactions) rather than the two of Cronbach and Kenny's SIRM. Although a full discussion of those factors is beyond the scope of this chapter, the three-way combination is particularly important to the study of stereotype inaccuracy because it tests for ingroup bias. The Subject Group × Target Group × Attribute factor tests whether stereotype exaggeration or underestimation (the attribute effect) is more likely to occur when people from group A judge people from group B and when people from group B judge people from group A than when people from group A judge people from group A and people from group B judge people from group B. If, for example, stereotype exaggeration only occurs when people judge groups other than their own, one would have an ingroup bias effect.

4. Components as a Means to Measure and Eliminate Noise In Judgments

Componential approaches are particularly useful for assessing and eliminating noise (judgment factors irrelevant to accuracy) from judgments. For example, consider a case in which both judgment and criteria are measured on a subjective scale, as might be the case when judging personality traits and the criteria are target self-reports. There could be a discrepancy between judgment and criteria not because perceivers are inaccurate but because they use the judgment scale in a different manner than do targets. Such subjective differences (or response biases) can be measured and eliminated when using a componential approach. Typically they manifest as one or more of the various elevation components in the different models (see, e.g., Kenny, 1994).
for a detailed discussion of how componential approaches can assess and remove response biases.

5. Must Components Be Assessed In All Accuracy Research?

Ever since Cronbach's (1955) review, researchers have emphasized the importance of assessing components, sometimes going as far as to claim or convey that components must be assessed to address accuracy questions (see, e.g., Hastie & Rasmussen, 1988; Jones, 1990; Judd & Park, 1993; Kenny & Albright, 1987). Whether all accuracy research must assess components, however, depends on the precise meaning of this claim. Nett, therefore, moderate and strong forms of this claim are considered.

a. The Moderate Form: Should Researchers Understand Existing Componential Approaches? The moderate form of this claim suggests that accuracy researchers should understand existing componential approaches to have better insights into the meaning of the results obtained from studies that do not explicitly assess components. This is undoubtedly true. It certainly behooves researchers interested in accuracy research to have more, rather than fewer, insights into the potential sources of social perception and the processes leading to accurate or inaccurate judgments and, especially, of the limitations and potential artifacts that influence whatever index of accuracy they do use.

b. The Strong Form: Process Versus Accuracy (Again). The strong form of the claim is that all accuracy researchers must explicitly perform componential analyses because otherwise their research will be meaningless or uninterpretable. A variety of logical, theoretical, and empirical considerations indicate that this strong claim is overstated.

First, componential approaches provide one class of explanations for how a person arrived at an accurate or inaccurate judgment. Indeed, Cronbach (1955) titled his article "Process Affecting Scores on 'Understanding of Others' and 'Assumed Similarity'" (emphasis mine). Why? Because components provide information about the processes of judgment.

How do I arrive at my prediction that Mike Piazza will hit a grand slam? Do I always predict that he will hit home runs, or am I a particularly astute judge of Mike's hitting? Why does Louise think African Americans are less likely to complete high school than they really are? Does she underestimate every group's likelihood of completing high school? Does everyone, including African Americans, underestimate African Americans' likelihood of completing high school? Or is Louise ethnocentric, underestimating only African Americans' success and not her own group's success at completing high school? Componential analyses provide answers to these types of questions by addressing the processes by which people arrive at accurate or inaccurate social judgments.

However, process is irrelevant with respect to establishing the degree of (in)accuracy of some perception. If I say "Mike is going to hit a home run" and he does, this particular prediction is right. End of discussion regarding my degree of accuracy. Although a single judgment is rarely psychologically interesting, the principle is just as true for multiple judgments. If I say that Mike Piazza is the best home-run-hitting catcher of the modern era and that he hits more home runs than any other catcher, again, I am right.

With respect to understanding how I arrived at that prediction, it would be valuable to estimate my elevation, stereotype accuracy, differential elevation, and differential accuracy (if you like Cronbach's or, if you prefer, SRM, my elevation, my perceiver effect, Mike's target effect, and our interaction effect. But if one wants to determine whether my prediction is accurate, the only thing we need to do is figure out whether he hit the ball over the outfield fence in fair territory (in the case of the single judgment), and whether he hit more home runs during his career than any other catcher (in the case of the test of multiple judgments).

If non-Armenians commonly believe that Armenians are public parasites burdening the financial community with their constant need for charity far more than other groups, and Armenians actually make fewer demands on public charity than other groups (LaPiere, 1936), then non-Armenians overestimate the financial burden created by Armenians. Again, period, the end, with respect to establishing the inaccuracy of this belief.

Interpreting that inaccuracy is another matter. As Ryan, Park, and Judd (1996) have pointed out, in the absence of their null accuracy componential design, we cannot conclude, as did LaPiere (1936), that this means those holding nasty beliefs about Armenians are necessarily anti-Armenian bigots. Perhaps they overestimate every group's need for charity, including that of their own. In that case, they are not ethnocentric at all. People with nasty beliefs about all groups including their own may be mean and evil, but they are not ethnocentric. Judd and Park's (1993) componential design would be extremely useful for providing some insight into why they overestimate Armenians' request for charity, but it is completely irrelevant with respect to establishing whether anyone overestimates Armenians' requests for charity. That question can only be answered by comparing estimates of Armenian need for charity to some criteria.

c. No Reflection of Components. It is extremely tempting to try componential approaches to accuracy. First, they are statistically rigorous and complex, which gives them a particularly scientific aura. Second, they capture important, fundamental aspects of social perceptual judgment processes. Third, they successfully identify sources of bias or noise in judgments
that few researchers usually mean by accuracy. Thus, it is tempting to view components as concrete features on the social perceptual landscape. If they are few, then should not accuracy researchers have to assess them?

Such absolute positions regarding components ("Cronbach’s [or SRM] components must always be assessed" or "accuracy can only be viewed componentally," etc.) are overrated for several reasons. First, there is no single right way to divide up components of social perception, as should be clear from the brief review of Cronbach’s, Kenny’s, and Judd and Park’s componential approaches. They have important similarities, but, obviously, there are also differences between all three. Such differences are made salient when the three approaches appear side-by-side, as they do in Table II. If there were any single "right" set of components that "must" be examined, if components were actually hard and fast fixtures in the social perception landscape, there could not possibly be three different breakdowns of components unless one breakdown in "right" and the other two are "wrong" or unless each was woefully incomplete.

If all were partially right, but incomplete in that they failed to address components identified by other researchers, then a full componential model would need to assess all the components identified by all models. Such a model is presented at the bottom of Table II. If components are "real" and "must" be assessed, then the only complete way to do it would be to assess the more than 50 components identified in this model. Such a model has never been recommended even by advocates of componential approaches and is not being recommended here. Indeed, it is an extreme as to border on absurd—but such an absurd model might be required if all components "must" be assessed.

The situation, however, is far more complex than even this hypothetical combined componential model suggests. There is a potentially infinite number of ways in which social perception could be broken down into components (see also Kraghamski, 1989). Attributes could be further broken down into a variety of types or subcategories (e.g., positive vs. negative; explanations vs. descriptions vs. predictions; behaviors vs. traits, and so on). Similarly, both perceivers and target groups could be broken down not only by ingroup and outgroup but by any of the infinite ways of identifying groups (culture, demographic characteristics, memberships in organizations, professional expertise, etc.).

This is not meant to suggest, however, that existing componential approaches are purely subjective and arbitrary and, therefore, can be ignored, but the choice of components will depend entirely on the types of processes one would be interested in studying and the types of response biases one would like to assess or eliminate. Different componential breakdowns serve different purposes and provide insights into different aspects and processes of social perception. Thus, understanding existing componential approaches would seem crucial to anyone studying accuracy to gain insights into how best to interpret their own or anyone else’s data addressing the degree of (in)accuracy in social perception.
may be lost by not performing a full componential analysis is also explicitly discussed.

1. Correlational Approaches

Most noncomponental approaches to assessing accuracy or processes underlying accurate and inaccurate social perceptions use correlations (typically Pearson's, occasionally regression coefficients) to assess the extent to which judgments correspond to criteria. In general, when judgments concern a single attribute, correlations between judgments and criteria capture Cronbach's (1955, p. 191) differential accuracy correlation, which he described as "sensitivity to individual differences." These are the only processes included in present measures of social perception which depend on [perceiver's] sensitivity to the particular D target."

The simplest and most typical form of correlation in accuracy research is that between a set of perceivers' judgments or predictions regarding a single trait or attribute of a set of targets and a criterion measure of that attribute or trait. For example, teachers' predictions regarding students' achievement may be compared with grades or standardized test scores; interviewers may evaluate a set of interviewee's skills, which are then compared with work samples; or perceivers may estimate the percentage of people belonging to various demographic groups that complete college, which are then compared to Census figures. Such correlations systematically remove the elevation and stereotype accuracy components from correspondence between perceivers' judgments and the criterion. (This is because these correlations reduce data to deviations from the mean.) Thus, a simple correlation (between judgment and criterion) goes a long way toward eliminating many of the biases, artifacts, and problems in assessing accuracy first identified by Cronbach.

Of course, the correlation coefficient is not perfect. First, it removes or avoids but does not directly assess Cronbach's (1955) elevation and stereotype accuracy components. Because correlations remove average differences between judgments and criterion, they cannot assess any consistent tendency to over- or underestimate targets (elevation & stereotype). If it were important to assess those components to address some research question, one could not use the correlation to do so.

Cronbach was, in addition to differential accuracy correlation, referring to differential elevation correlation in this quotation, which refers to the correlation between the perceivers' judgment of a target averaging over all attributes and the targets' average score on the criteria encompassing those attributes. This, however, is irrelevant when there is only a single attribute being judged.
Second, correlations equate the variability of judgments and criterion. Therefore, they cannot assess whether perceivers consistently over- or underestimate target variability.

Because mean and variability differences between judgments and criteria probably often reflect response bias or scaling discrepancies between perceivers and criteria, these limitations to correlations do not greatly under- mine their utility in assessing accuracy. The term "scaling discrepancies" is used here to refer to the idea that people may use scale points in a manner differently than used in the criterion. This would obviously be true if, for example, judgment and criterion are assessed in different metrics (e.g., subjective rating scale and percentages, respectively).

People, however, still might use the numbers in some scale differently than is used for the criterion, even if they are supposedly on the same scale. For example, let's say Alfred estimates three people's IQ scores as 40, 50, and 60, when they are really 115, 120, and 125. Although it is possible that Alfred believes three of these fairly intelligent people are neurologically impaired, it is more likely that Alfred does not fully understand how IQ scores are scaled. He dramatically underestimates people's IQ in absolute terms, but his estimates are also overly sensitive to actual variations in IQ (Alfred's judgments go up 10 IQ points for every 2-point increase in actual IQ). Given his subjective IQ scale, however, the correlation between his judgments and actual IQ would be perfect (1.0), because mean differences in judgment criteria are irrelevant to computation of the correlation, the correlation coefficient is computed after statistically equating the variability in judgment and criterion, and his judgments move in (differently scaled) lockstep with target's actual IQ.

Thus, the correlation coefficient would yield a conclusion that Alfred is an excellent judge of people's intelligence. Is the conclusion justified? As long as one keeps in mind that what this really means is that "Alfred is very good at detecting differences in people's intelligence, but this does not tell us anything about whether he consistently over- or underestimates people's intelligence," it is perfectly justified. 8

2. Construct Validity and Correlational Approaches To Accuracy

The section on criteria argued that assessing accuracy was much like assessing the validity of many social science constructs. This is also important here, because the correlation coefficient is so frequently used to establish the validity of some measure that it is often referred to as the "validity" or "validity coefficient" of some measure (e.g., Campbell & Stanley, 1963; Cook & Campbell, 1979; Dawes, 1979). In much the same manner, correlations can be used to assess the accuracy of social perception.

Establishing accuracy, however, is somewhat more complex than establishing construct validity because social perception, judgment, and expectations are themselves constructs. Thus, all the rules involved in establishing construct validity arise not just when measuring targets' attributes but when measuring perceivers' expectancies, perceptions, judgments, and beliefs about others. Accuracy, therefore, is not usually best reflected by correlations between observable measures (e.g., a measure of perceiver expectations and a measure of target extraversion). Accuracy will often be reflected by correlations between the underlying constructs representing the social perception and the criteria (these often be accomplished either by disattenuating correlations for unreliability or by using LISREL-type models; see, e.g., Bollen, 1989; Carmines & Zeller, 1979).

This should not, however, be misinterpreted to mean that all accuracy research must necessarily assess relationships between underlying constructs rather than observed measures. Sometimes it may just not be possible to do so, as, for example, when there is only a single item assessing either the perception or the criterion. This, of course, does not constitute any sort of immovable obstacle to or fatal flaw in accuracy research. Correlations among such measures, however, will tend to underestimate accuracy to the extent that the observed measures imperfectly reflect the underlying attributes or expectations. Therefore, people may be more accurate than indicated by research that only assesses correlations between observed measures of expectations and observed measures of criteria.

This brief delving into construct validity, correlations, and unmeasured variables was necessary to lay the foundation for understanding three of the main noncompositional approaches for assessing accuracy in social perception. All three are fundamentally based on the correlation between perception and criteria.

3. Brunswik's Lens Model

Brunswik (1952) suggested that accurate perception of reality (both object and social) involves the use of cues probabilistically related to objective reality. He metaphorically called his approach the Lens Model to capture the idea that objective reality is never observed directly. Instead, cues related to objective reality must be observed and interpreted as relevant to some judgment; that is, objective cues are seen through the "lens" of subjective perception and judgment.
This does not necessarily mean that perception is a purely subjective phenomenon unrelated to objective reality. Indeed, one of the main purposes of the Lens Model is to provide a mechanism not only for assessing people’s degree of accuracy but also for understanding sources of both accuracy and inaccuracy in their judgments. The Lens Model has been used to study accuracy, inaccuracy, and bias in a wide variety of domains, including education, criminal, and interpersonal contexts. The pattern of the arrows represents some sort of psychological construct that cannot be directly observed (self-esteem, extraversion, intelligence, etc.). The Cues, shown in the middle of the figure, are directly observable or measurable phenomena. The arrows pointing from the Psychological Attributes to the Cues are labeled “Validity,” because they represent the extent to which the underlying attribute manifests itself in the observable Cues.

The third arrow represents perceivers’ judgments (or perceptions). The arrows going from the Cues to Judgments represent the extent to which those observable cues influence perceivers’ judgments (labeled Cue Utilization). The Lens Model characterizes social perception as a two-step process: observable manifestation of psychological attributes and perceivers' use of observable cues to arrive at judgments. Accuracy, therefore, is captured by the correlation between the psychological attributes and the judgments (the long, double-headed arrow in Fig. 1). Correlations assess how well the judgments correspond to the attributes—that is, accuracy.

The Lens Model is a noncompositional, correlational model for assessing both degree of accuracy and processes of social perception. Identifying cue validity and cue utilization focuses on a very different set of the processes than is typically the focus in compositional models. As such, it provides different (but not better or worse) types of insights into processes of social perception than do compositional models.

4. The Realistic Accuracy Model

Funder’s (1995, 1999) Realistic Accuracy Model (RAM) draws on the set of fundamental assumptions described under “probabilistic realism” to create a model that could be viewed as an extension and elaboration of Brunswik’s Lens Model. Some of the main ideas of RAM are depicted in Fig. 2. As with the Lens Model, overall accuracy is typically assessed by the correlation of the underlying attribute with the perceivers’ judgment (represented by the large, curved, double arrow on the bottom of Fig. 2). Four steps are needed for perceivers to arrive at an accurate judgment: displayed in between the underlying attribute and the judgment.

First, the underlying attribute needs to create some sort of observable evidence related to that attribute (the cues, in the Lens Model). Dissonance, for example, is not likely to be displayed in a large lecture hall (except perhaps during test time). Interest in the class is more likely than honesty.

![Diagram](image-url)
overall accuracy is typically assessed via correlations, although discrepancy scores (between judgment and criteria) can be used, too (see Funder, 1987, 1995). RAM is particularly good at explaining why accuracy in person perception may often be low. For the judgment to closely correspond to the criterion, that criterion needs to clearly manifest itself in ways that could be and, in fact, are detected by the perceiver, and then the perceiver needs to use that detected information (as well as not use information that is not relevant to the judgment). A breakdown at any step will dramatically undermine accuracy. Furthermore, by focusing on combinations of perceiver, target, attributes, and evidence, RAM is also particularly good at highlighting processes that may enhance or undermine accuracy.

5. Dawes's Improper Linear Models

Dawes (1979) made a very interesting discovery. In reviewing his own and others' research on decision-making, he found that people tend to be very good at identifying the evidence or cues that are relevant to making some prediction but are not very good at combining or integrating those cues. Thus, their overall predictive accuracy tends to be quite low. Note, however, that this is not because people are completely inept. They are good at one part of the prediction task (identifying criteria for making a prediction) but poor at another part (putting those criteria together).

One of Dawes's (1979) examples involved admissions to graduate school in psychology. The criteria typically used for making admissions decisions seem appropriate: GRE scores (general intellectual ability), GPA (achievement at academic tasks over an extended period), and letters of recommendation (what experts in the field who are highly familiar with the applicant have to say about him or her). Nonetheless, Dawes (1979) found that the correlation of graduate admissions committee evaluations with later success in graduate school is typically quite low (.16).

If people were completely inept, they would not even use appropriate criteria; that is, the criteria they do use would not predict success in graduate school. However, if they are good at identifying the appropriate criteria, but use them poorly, then the raw criteria themselves should do a much better job at predicting graduate success. This was indeed the case—the overall multiple correlation of the criteria with graduate success was about .4.

What to do? It is unreasonable to expect admissions committees to compute complex statistical formulas in their heads or to create a formal statistical score for each applicant. Dawes provided an elegant and simple way of making sense of the situation that might lead to better results. Dawes provided a way of using the evidence or cues that are relevant to making a prediction (as in going to a statistician) and a way of using the evidence or cues that are relevant to combining or integrating those cues (as in combining or integrating those cues).
GPA, and excellent letters of recommendation might receive 2-scores of 1, 2, and 2, respectively, for a total score of 5. George, with high ORs, a good GPA, and good letters, might receive 2-scores of 2, 1, and 1, for a total of 4.

Pricilla would be ranked higher than George.

This is different from a formal statistical model primarily because the weights for each predictor have been chosen in a less than optimal manner (many statistical prediction techniques, such as regression, identify how to weight the criteria in such a manner as to maximize their overall predictive validity). Dawes's (1979) analysis showed, however, that this "unoptimal" weighting of criteria predicts outcomes nearly as well as do formal statistical models. In the graduate admissions example, Dawes' improper linear model correlated .38 with future success in graduate school. Dawes (1979) went on to show that a simple, improper linear model performed similarly well in predicting all sorts of outcomes, including choice of bullet-type for a police department and a bank's predictions regarding companies likely to go bankrupt.

Dawes' improper linear model is fundamentally different than the Lens Model and RAM. The Lens Model and RAM were specifically designed to assess degree of accuracy and processes underlying social perception. In contrast, Dawes' model is primarily descriptive (it suggests how people should go about making decisions and arriving at predictions).

Nonetheless, it has been included here for two reasons. First, Dawes' (1979) conclusion that people are good at selecting criteria, but not good at using them, is descriptive. In RAM terms, it indicates that people often are good at detecting available and relevant cues, but that they often do not use them well (as Lens Model terms, their cue utilization would be poor). Second, although Dawes did document that people were, on their own, not very good at arriving at accurate predictions, he also showed that the accuracy of their predictions could easily be improved. Identify the criteria, weight them equally, and add!

C. COMPONENTIAL AND NONCOMPONENTIAL APPROACHES TO THE STUDY OF ACCURACY

Componential approaches to the study of accuracy have provided several major contributions to the study of social perception. As a research tool for assessing the processes of social perception, they demonstrate again how

Dawes's model was actually even more accurate. The actual model produced linear model generally (i.e., not just in the graduate admissions example) predicted outcomes better than the split sample cross-validated regression weights.

ACCUACY IN SOCIAL PERCEPTION

accuracy research typically provides the type of double contribution that few, if any, experimental studies focusing exclusively on error, bias, and process can provide. Componential approaches address accuracy, inaccuracy, and process. They are particularly useful at identifying some of the reasons a perceivers is a particularly good or poor social judge and also can demonstrate that some perceivers might be accurate in some ways, yet inaccurate in others, even when judging the same target. Last, they are especially good at measuring and eliminating response biases that are largely or completely irrelevent to the assessment of accuracy.

It is also clear, however, that componential approaches have not been the only way to productively study accuracy. Three noncomponential approaches, Brunswik's (1952) Lens Model, Funder's (1995) RAM, and Dawes' (1979) improper linear models were also discussed. All rely on the correlation as the primary means for establishing correspondence between social judgment and criteria, and all have provided important insights into both the inaccuracy of social perception and basic social perceptual processes. Other approaches relying on Bayer's theorem (e.g., Krueger, 1996; Krueger et al., 2003; McCauley & Stitt, 1978; McCauley et al, 1980) and structural equation modeling (e.g., Jensen, 1991; West & Anderson, 1976; Williams, 1976) have also contributed to understanding processes leading to accurate and inaccurate social perceptions, especially with regard to interpersonal expectations and social stereotypes.

It is clear, therefore, that although componential approaches are valuable and important, one need not need perform any particular componential analysis to assess accuracy. Indeed, in the absence of data permitting a full componential analysis, performing a simple correlation between belief and criterion, though imperfect, would typically provide far more information than would doing nothing. Even in such a situation, however, it behoives the researcher performing the correlation to understand componential approaches to appropriately interpret just what that correlation does and does not show.

IX. Conclusions

This review has documented many of the conceptual, methodological, and statistical advances in the psychological study of social perceptual accuracy, with a particular focus on addressing many of the criticisms of accuracy research that have been raised over the last 50 years. In doing so, the conclusion has been repeatedly reached that although there are genuine difficulties, ambiguities, or obstacles to accuracy research, they are no greater than in most other areas of social psychological research. Indeed, this
review has shown that the conceptual, theoretical, methodological, and empirical bases of accuracy research are identical to those of much other research in psychology. By taking stock of the many conceptual and methodological advances in accuracy research that have been made over the last 20 years, this review has attempted to provide a road map for further accuracy research. Although it is impossible for such a map to represent every twist and turn in the road, it has attempted to provide some valuable signposts providing direction in the pursuit of accuracy research. These signposts show how to circumvent or avoid many known obstacles, indicate that the road is clearer in many directions that were once believed to be filled with obstacles, and have identified many alternative routes available when known obstacles arise.

A. NONE OF THE COMMON CRITICISMS OF ACCURACY RESEARCH THREATENS ITS VIABILITY

This review has contested the viability of some of the common theoretical, conceptual, and methodological criticisms of accuracy research. Some criticisms were seen to be manifestly false; others were seen to be logically incoherent. Yet others were seen to be partially valid but not to constitute serious threats to accuracy research. This review also showed that even completely valid criticisms of accuracy research could readily be addressed by a clear understanding of its nature and, limitations to, the insights into social perception that accuracy research can provide.

B. THE CRITERIA FOR ACCURACY ARE IDENTICAL TO CRITERIA USED IN PSYCHOLOGICAL, SOCIAL, AND BEHAVIORAL SCIENCE RESEARCH

The crucial issue of identifying criteria for establishing accuracy was also explored. This review contested arguments criticizing accuracy research for alleged difficulties in identifying appropriate criteria for validating social beliefs. Such criticism would seem to logically require a similar criticism of nearly all research in psychology (with the possible exception of biological research). The logic of establishing accuracy of social perception overlaps almost completely with the (far less controversial) logic of establishing construct validity in the social sciences.

C. BOTH COMPONENTIAL AND NONCOMPONENTIAL APPROACHES TO ACCURACY ARE VALUABLE

Furthermore, several major approaches to the assessment of accuracy and processes leading to accurate and inaccurate social judgments were reviewed. This included reviews of the three main componential approaches to accuracy and three of the major noncomponential approaches to the study of accuracy. It was concluded that both componential and noncomponential approaches have made major and significant contributions to understanding social perception and accuracy.

D. PROCESS RESEARCH ON ERRORS, FLAWS, AND BIAS IS NOT MUTUALLY EXCLUSIVE WITH ACCURACY

Despite the often-demonstrated existence of a slew of logical flaws and systematic biases in lay judgment and social perception, such as the fundamental attribution error, false consensus, over-reliance on imperfect heuristics, self-serving biases, and so forth, people's perceptions of one another are surprisingly (though rarely perfectly) accurate. Despite being most well-known for their self-fulfilling effects, teacher expectations, for example, typically predict student achievement more because they are accurate than because of expectancy effects (e.g., Brophy & Good, 1974; Brophy, 1983; Eccles & Wigfield, 1985; Jussim, 1991; West & Anderson, 1976). People are also often surprisingly good at judging one another's personalities (e.g., Funder, 1987, 1999; Funder & West, 1993; Paunonen, 1991). People in close relationships are often better than strangers at inferring each other's thoughts and feelings (Ickes, 1993, 1997; Struik & Ickes, 1992). Indeed, people often know more about a partner or a friend than they do about a stranger (Markus & Zajonc, 1972; Zeisel, 1976). People in close relationships are often better than strangers at inferring each other's thoughts and feelings (Ickes, 1993, 1997; Struik & Ickes, 1992). Indeed, people often know more about a partner or a friend than they do about a stranger (Markus & Zajonc, 1972; Zeisel, 1976).

E. BELIEFS ABOUT SOCIAL GROUPS (STEREOTYPES) OFTEN SHOW MODERATE TO HIGH ACCURACY

Perhaps most surprisingly, considerable research has accumulated showing that people's perceptions of groups and individual targets from differing race, sex, and social class backgrounds are often quite (though
not perfectly) accurate (Dickman et al., 2002; Hull & Carter, 1999; Judd et al., 1993; Jussim et al., 1996; Lee et al., 1995; Krueger, 1996; Madson et al., 1996; McCauley & Stitt, 1978; Ryan, 1996, 2002; Ryan & Bogat, 2001; Swim, 1994; Williams, 1976). Because, however, accuracy is not mutually exclusive with, and sometimes is even enhanced by, bias, this claim does not refute or deny the existence of prejudice or discrimination. Indeed, many of the same studies demonstrating moderate to high accuracy have also demonstrated that stereotypes do sometimes lead to systematic biases in judgments of groups and individuals.

F. UNDERSTANDING ACCURACY IS ESSENTIAL FOR UNDERSTANDING AND ALLEVIATING SOME SOCIAL PROBLEMS

Research on accuracy would appear essential to the application of psychology to the remediation of many social problems, especially those involving intergroup relations. Beliefs about groups have for decades been presented as either a cause of social inequities or as a reflection of prejudice and existing dominance hierarchies, or both (e.g., Allport, 1958; American Psychological Association, 1991; Fiske & Taylor, 1991; Iost & Banaji, 1994; LaPiere, 1936; Picketing, 2001; Sidanius & Pratto, 1999). If such beliefs are largely erroneous, and if they cause injustices, as many such perspectives imply, then the solution to many group-based social problems should focus almost entirely on correcting the beliefs and perceptions of those holding such beliefs. If the problem is erroneous perception, then correcting the perceptions should eliminate, or at least greatly alleviate, that problem.

To the extent, however, that some beliefs about group differences are accurate because groups really do differ on some objectively valued attributes, then correcting erroneous perceptions would be a misguided endeavor, or at best a woefully incomplete solution. In this situation, efforts need to be focused on redressing the underlyling inequalities, which will probably involve devoting resources to improving the status of lower status or unjustifiably stigmatized groups. Accuracy research, therefore, would appear to be crucial with respect to identifying how to devote resources to the alleviation of social problems—with relatively more resources being directed to correcting perceiver beliefs when those are shown to be inaccurate but relatively more resources being directed to improving the status of low status groups when perceiver beliefs are shown to be accurate.

G. AN INVITATION TO STUDY ACCURACY

There has been a great blossoming of accuracy research over the last 20 years (Ambady & Rosenthal, 1992; Funder, 1987, 1995, 1999; Ikeda, 1993, 1997; Kenney, 1994; Kenney & Albright, 1987; Judd & Park, 1993; Jussim, 1989, 1991; Jussim et al., 1996; Lee et al., 1995; Oakey et al., 1994; Ryan, 2002; Swim, 1994). Now that the major obstacles to accuracy research have been cleared away and that new and innovative ways to address accuracy are being continuously developed, if anything, the pace of research on accuracy should accelerate over the next 20 years.

Nonetheless, specific future developments are difficult to anticipate. Twenty years ago, I would never have predicted that people could be as accurate as they apparently are on the basis of minimal information (Ambady & Rosenthal, 1992) or that stereotypes would show so much accuracy as they consistently seem to show (Brodet & Rose, 1998; Hall & Carter, 1999; Lee et al., 1995, McCauley & Stitt, 1978; McCauley, 1995; Ryan, 2002; Swim, 1994, Wohlschlag et al., 2000). Nonetheless, many classic issues and recent developments in social psychology naturally raise a host of interesting and important questions about accuracy.

1. Moderators: The Person, the Situation, and the Relationship

Social and personality psychology have a long tradition of examining the role of individual differences, situational forces, or both in causing or moderating all sorts of social phenomena. This tradition naturally lends itself to the study of accuracy. Hull and Carter (1999), for example, examined whether a slew of individual difference factors increased or reduced the accuracy of people's beliefs about differences between men and women (the most successful moderator was nonverbal sensitivity). With the modern methods that have cleared away the old obstacles, we may finally be able to obtain some answers to the question that dominated early accuracy research and seemed to have no answer: 'Who is a more accurate judge of others?'

Similarly, researchers have also begun addressing the other classic social psychological issue—situational moderators of accuracy. Wocnik et al. (1999), for example, found that White students, given a "multicultural" set (one emphasizing the importance of understanding and valuing bona fide racial and cultural differences), held more accurate racial stereotypes than those given a "color-blind" set (one emphasizing the sameness of all people and deemphasizing the importance of group differences). Those given
the multicultural set also used the stereotypes more appropriately when judging individual targets. Thus, situations evident different sociopolitical mindsets moderated stereotype accuracy.

Accuracies may also vary by aspect of relationships. Length of relationship seems like an obvious moderator (the longer people get to know one another, the more accurately they should see one another). Although research on stereotypes is broadly consistent with this hypothesis (especially the overwhelming evidence that people rely less on stereotypes the more individualizing information they have; Eagle, Makhlajian, Ashmore, & Longo, 1991; Kunda & Thagard, 1996), research using the S&BM has not always found that accuracy increases with length of relationship (Kenny, 1994). Whether accuracy also systematically varies by type of relationship (e.g., romantic, parent-child, occupational, etc.) is currently unknown.

2. Accuracy and Bias and Their Integration

Whether lay social beliefs and perceptions are typically rational and accurate or inaccurate and biased has been controversial for decades. Clearly, however, those emphasizing error, bias, and the ways in which social beliefs create social reality have dominated the literature on social cognition (e.g., Fiske, 1998; Janis, 1982; Kahneman & Tversky, 1973; Nisbett & Ross, 1980; Snyder, 1984). These views have created an image of a social perceiver whose misbegotten beliefs and flawed processes construct not only illusions of social reality in the perceiver’s own mind but actual social reality through processes such as self-fulfilling prophecies. In this bleak view, the mind becomes primarily a product of cognitive shortcomings and distorted social interactions.

Implications of the prevalence of fallacies to the power of error and bias is that something important has been discovered. Because error, bias, and accuracy can coexist simultaneously, however, merely demonstrating error or bias does not demonstrate inaccuracy. Although people undoubtedly commit errors and biases and are rarely perfectly accurate, almost none of the literature routinely cited as substantiation to the power and prevalence of error and bias actually tests for accuracy. As a consequence, despite the manner in which it is cited, that literature provides little direct information about accuracy. One has to actually test for accuracy to reach conclusions about inaccuracy.

Putting aside some of the huge existing literature on error and bias into some context by also testing for accuracy may be less difficult than it seems, because methods are being developed for extracting information about accuracy, agreement, and objectivity from studies that have focused exclusively on inaccuracy, disagreement, and subjectivity (Jussim, 2004; Jussim, Harber, Crawford, Cain, & Cohen, in press). Such methods suggest, for example, that classic demonstrations of bias and subjectivity, such as Rosenhan (1973) and Hastorf and Castelli (1954), actually provide far more evidence of accuracy and objectivity than of bias or subjectivity. At least under some conditions, furthermore, heuristics, biases, and accuracy may be integrated. Many of the classic tenets on flaws and biases (e.g., Kahneman & Tversky, 1973; Nisbett & Ross, 1980) include disclaimers and phrases along the following lines: The many biases, flawed processes, and faulty heuristics we have identified may often serve people quite well. Although the error and bias classics rarely expanded much on this type of claim to show how, why, or when it might actually be correct, the claim itself is correct, because a variety of reasons documented here bias will sometimes increase accuracy, a flawed or imperfect process may still be good enough to generate mostly accurate predictions or judgments, and expert models are sometimes contradictory, implying at least one is wrong or not applicable, so that demonstrating that people deviate from such models does not always mean they are inaccurate.

Both theory and empirical research have begun to address the role of heuristics and biases in increasing accuracy (Dawes, 1988; Jussim, 1991; Kenny & Aciello, 2001). Indeed, it seems likely that the fact that many heuristics and biases do increase accuracy will help explain and develop and are difficult to eliminate. As such, a deeper understanding of the social cognitive processes underlying judgment, prediction, and decision-making will undoubtedly be obtained by research well beyond the benefits, and not just the costs, of well-documented biases and heuristics.

3. Accuracy of Implicit Beliefs

The type of reaction time, categorization, and priming techniques common to research on implicit and automatic social cognition (Bergh & Charrand, 1999; Fazio, Jackson, Dunton, & Williams, 1995; Grenewald, McGhee, & Schwartz, 1998) at first glance may seem to readily lend themselves to the study of accuracy. It is not meaningful, for example, to discuss the “accuracy” of a reaction time. Furthermore, much of this type of research has focused on attitudes, and attitudes cannot be accurate or inaccurate.

Nonetheless, research on the accuracy of implicit and automatic beliefs can and should be conducted. One key question involves the extent to which implicit and automatic beliefs correspond to social reality. For example, do...
people's implicit beliefs about differences between groups correspond to the size of the actual group differences? To address this type of question, people's scores on the measures reflecting their implicit beliefs (priming effects, implicit associations, etc.) can be correlated with criteria regarding those beliefs. For example, implicit beliefs about differences between African Americans and Whites can be correlated with Census reports, or implicit beliefs about a particular person can be correlated with that person's attributes. Much of the work to date on implicit beliefs and attitudes emphasizes their dark side, in large part, because it has often focused on prejudice (Fazio et al., 1995; Greenwald et al., 1988). This emphasis is well-justified by the damage that work to date on implicit beliefs and attitudes inflicts on individuals and groups and is broadly consistent with social psychology's historical emphasis on error and bias. Nonetheless, the extent to which implicit beliefs about the self, other individuals, and groups reflect social reality (in addition to, and, perhaps, sometimes rather than) perceiving errors and biases is currently unknown.

4. Evaluating the Effectiveness of Interventions Designed to Increase Accuracy

Accuracy interventions, although they are rarely referred to in those terms, pervade everyday life:

- Psychotherapy: The goals of personal insight, recognition of dysfunctional behavior patterns and ways to avoid them, and the reframing of unjustifiably blaming views of self or others are often goals of psychotherapy.
- Education: A liberal arts background, especially with experience in the humanities and social sciences, can be viewed as, in part, one immense undertaking intended to dismiss students of their inaccurate and parochial views of themselves, their peers, their country, and the various religions, cultures, nationalities, and ethnicities of the world. Advanced instruction in social, organizational, and cognitive psychology: business programs: statistics: and philosophy often provides specific training relevant to eliminating many of the errors and biases that are so well documented in the psychological literature.
- Multiculturalism/diversity: Interventions in both business and educational settings designed to reduce prejudice and intergroup conflict by increasing support for multiculturalism and diversity have become commonplace. Multiculturalism/diversity typically has a moral component ("we should value diversity") that is beyond the scope of this chapter, but multiculturalism also has an accuracy component ("we should replace our false beliefs about other groups with more accurate ones"). This latter point is rarely put so bluntly, but law can we be "sensitive to others' cultural differences" if we do not have some valid insights into and understandings of those differences? This type of point is frequently made in pleas for people to become more multiculturally sensitive. For example, Pinderhughes (1989) titled her book, which is essentially an extended treatise advocating multiculturality in therapy, Understanding Race, Ethnicity, and Power. Presumably, she believes that such understanding can be attained, otherwise she would not have written the book or given it that title.

From the perspective here, this constitutes a call for an increase in the accuracy of people's beliefs about (understanding of, insight into) those from other cultural backgrounds. I would call that an increase in the accuracy of their stereotypes, because I simply define a stereotype as a belief about a group. Indeed, Pinderhughes (1989, p. 147) makes essentially the same point when she identifies several abilities that enhance people's multiculturally competent, two of which are "The ability to control, and even change false beliefs, assumptions, and stereotypes" and "The ability to respect and appreciate the values, beliefs, and practices of persons who are culturally different.

Respecting others' values, beliefs, and practices makes sense primarily if one has a reasonably clear (accurate) sense of what those values, beliefs, and practices are.

Unfortunately, however, the accuracy of the new beliefs created by interventions such as therapy, education, and multicultural programs is rarely tested. The validity and effectiveness of psychotherapy remains controversial (Dawes, 1994; Seligman, 1995). The specific issue of whether clients' beliefs typically become more accurate through therapy is largely unknown. As a consequence, this would appear to be an interesting and important area for future research.

Similarly, little research has evaluated the extent to which education improves accuracy. On the one hand, direct explicit statistical training increases people's ability to use appropriate statistical reasoning when solving problems (Fong & Nisbett, 1991). On the other, the overwhelming majority of studies demonstrating biased judgments, overconfidence on faulty heuristics, and a myriad of failures to understand or apply principles of statistics or logic in everyday reasoning problems have relied on college student samples (e.g., Kahneman et al., 1982; Nisbett & Ross, 1980). Whether people who have received a college education generally reach more accurate conclusions about their friends, coworkers, and peers and about the larger groups in society than do their peers who have not gone to college but who are otherwise similar in age and intelligence is unknown. Presumably, the thousands of diversity and multiculturalism workshops, classes, and programs that have been instituted on campuses around the
country do achieve some degree of reduction in discrimination, if only because they convey the message that "people in power around here want nothing at all to get along." Furthermore, they are probably superior to doing nothing at all, but this is not guaranteed. Potential negative side effects, such as a dampening of the free exchange of views crucial to both a college campus and a functioning democracy, and an increase in the willingness to accept immoral or illegal behavior (e.g., acceptance of sexual harassment because someone is from a culture emphasizing machismo) "because that's how they do things in that culture," must be evaluated alongside any reductions in prejudice or discrimination. Nonetheless, the effectiveness of such programs has almost never been evaluated (e.g., McCloskey, Wright, & Harris, 2000). Thus, their effectiveness in leading people to become more acutely aware of (i.e., to more accurately perceive) how the beliefs, values, and practices of people from various backgrounds differ from their own is both unknown and a potentially rich area for future research.

H. "THE COMMON AND MAINLY VERIDICAL CHARACTER OF THE BASIC HUMAN PERCEPTIONS"

This review has suggested that it is no longer justified to dismiss this blooming of accuracy research for reasons such as accuracy research provides little information about psychological processes (process is a central focus of most accuracy research); research on accuracy is unnecessary because research on social cognition already shows that social perception is generally inaccurate (social cognition research has uncovered a host of errors and biases, but these do not demonstrate that people are generally inaccurate); research on accuracy often has not addressed how perceivers explain others' behaviors and attributes (it often has not, and this is a limit, but not a threat, to existing accuracy research); self-fulfilling prophecies may account for that which is "accurately" perceived (they may, but this does not preclude the study of accuracy); research on stereotype accuracy is unnecessary because even a stereotype that fits a group will inaccurately describe most members of that group (because this claim confounds levels of analysis, its challenge to stereotype accuracy research fails); it resists individual and group differences (existing scholarship on accuracy has never claimed that accurately perceived differences constitute immutable characteristics); it is hopelessly politicized (it sometimes evokes hostile, politically motivated criticism); and the complex in erelationships of accuracy, bias, and discrimination have received little discussion in the scholarly literature or even in everyday cultural discourse, but the research itself is probably no more politicized than many other areas of social psychology; it exacerbates social problems (it is a powerful tool for the redress of many social problems); criteria are so fundamentally flawed or ambiguous as to undermine the viability of accuracy research (the criteria are identical to those used to assess self-fulfilling prophecies and to assess validity of most psychological constructs); or much accuracy research has failed to perform a compensatory analysis (much has, and even when it has not, accuracy has been productively studied anyway).

Fifty years ago, Floyd Allport (1955), in responding to the New Look in Perception research which, arguably, launched the now-five-decade-long social psychological emphasis on error and bias, made the following appeal:

"What we are urging here is that social psychologists, in building their theories of perception, assume their share of the responsibility for reconsidering and integrating their 'social-perceptual' concepts, fraught with all their deviations and special cognitive loadings, with the common and mainly veridical character of the basic human perceptions" (p. 372).

It now seems clear that the meta-theoretical, theoretical, and empirical paths have been sufficiently cleared of obstructions for accuracy research to provide the balance to psychological perspectives on social perception that Floyd Allport called for 50 years ago.

Acknowledgments

I should thank Richard Aschener, Grendah Chapman, David Rinder, Joanne Kuege, and Clark McCloskey, for comments on earlier drafts of this chapter. This chapter is based, in part, on sections from a book by Lee Ross, currently in progress and tentatively titled Social Cognition and Social Realities.

References

Accuracy in Social Perception


