Profiting From Controversy

Lessons From the Person–Situation Debate

Douglas T. Kenrick — *Arizona State University*
David C. Funder — *University of Illinois*

ABSTRACT: For the past two decades the person–situation debate has dominated personality psychology and had important repercussions in clinical, social, and organizational psychology. This controversy strikes to the heart of each of these disciplines because it puts on trial the central assumption that internal dispositions have an important influence on behavior. According to emerging views of scientific progress, controversy serves the useful function of narrowing the field of competing hypotheses. In this light, we examine seven hypotheses that arose during the course of the person–situation debate, ranging from most to least pessimistic about the existence of consensual, discriminative personality traits. The accumulated evidence fails to support the hypotheses that personality traits are simply (a) in the eye of the beholder, (b) semantic illusions, (c) artifacts of base-rate accuracy, (d) artifacts of shared stereotypes, (e) due to discussion between observers (who ignore behavior in favor of verbal self-presentation or reputation), or (f) mere by-products of situational consistencies. Evidence also fails to support the hypothesis (g) that although traits are related to behavior, the relationship is too small to be important. Yet we have not simply come full circle to a reacceptance of traits as they were understood 20 years ago. Research generated by these hypotheses has allowed us to better specify the circumstances under which personality assessments will be valid.

Whether we are acting as professional psychologists, as academic psychologists, or simply as lay psychologists engaging in everyday gossip, the assumption that people have "traits" (or enduring cross-situational consistencies in their behavior) provides a basis for many of our decisions. When a clinical or counseling psychologist uses a standard assessment battery, he or she assumes that there is some degree of trait-like consistency in pathological behavior to be measured. When an organizational psychologist designs a personnel selection procedure, he or she assumes that consistent individual differences between the applicants are there to be found. When an academic psychologist teaches a course in personality, he or she must either assume some consistency in behavior or else face a bit of existential absurdity for at least three hours a week. Likewise, a good portion of our courses on clinical and developmental psychology would be unimaginable unless we assumed some cross-situational consistency. Even in everyday lay psychology, our attempts to analyze the behaviors of our friends, relatives, and co-workers are riddled with assumptions about personality traits.

Despite the wide appeal of the trait assumption, personality psychologists have been entangled for some time in a debate about whether it might be based more on illusion than reality (e.g., Alker, 1972; Allport, 1966; Argyle & Little, 1972; Bem, 1972; Block, 1968, 1977; Bowers, 1973; Epstein, 1977, 1979, 1980; Fiske, 1974; Gormly & Edelberg, 1974; Hogan, DeSoto, & Solano, 1977; Hunt, 1965; Magnusson & Endler, 1977; Mischel, 1968, 1983; West, 1983). Murmurs of the current debate could be heard more than 40 years ago (Ichheisser, 1943), but the volume increased markedly after Mischel's (1968) critique, and things have not quieted down yet (Bem, 1983; Epstein, 1983; Funder, 1983; Kenrick, 1986; Mischel, 1983; Mischel & Peake, 1982, 1983). Of late, discussants have begun to express yearning to end what some see as an endless cycle of repeating the same arguments. Mischel and Peake (1982) and Bem (1983), for instance, both use the term *déjà vu* in the titles of recent contributions, suggesting that they feel as if they have been here before. Other commentators maintain that the debate has been a "pseudo-controversy" (Carlson, 1984; Endler, 1973) that never should have occurred in the first place.

However fatiguing it may now seem to some of its erstwhile protagonists, the debate over the alleged inconsistency of personality has been more than an exercise in sophistry. In the course of the nearly two decades since Mischel's (1968) critique, a number of provocative hypotheses have been put forward, along with a host of studies to evaluate them. Platt (1964) and Popper (1959), among others, maintained that science typically progresses through the accumulation of negative information—that is, by eliminating hypotheses that data suggest are no longer tenable. From this perspective, it may be worth taking a look back at the hypotheses suggested during the consistency controversy, this time in the improved light shed by two decades of research. In this light, the debate can be seen as an intellectually stimulating chapter in the history of the discipline, replete with useful lessons for professionals who include assessment in their repertoire.

The "Pure Trait" Model and Its Alternatives

Discussions of the "person versus situation" debate traditionally begin with the "pure trait" model (Alston, 1975;

 0003-066X/88/$00.75
Vol. 43, No. 1, 23–34

Argyle & Little, 1972; Mischel, 1968): that people show powerful, unmodulated consistencies in their behavior across time and diverse situations. This position has been attacked frequently over the years. However, it is really just a "straw man," and even traditional personality researchers find it unacceptable (see, e.g., Allport, 1931, 1966; Block, 1977; Hogan et al., 1977; Jackson, 1983; Wiggins, 1973; Zuroff, 1986). Complete invariance in behavior is associated more with severe psychopathology than with "normal" behavior.

If the consensus rejects the "pure trait" position, then what can replace it? Several alternative hypotheses have been advanced over the years. These hypotheses differ with regard to four issues, which can be arranged into a logical hierarchy:

1. Consensus versus solipsism. Are traits merely idiosyncratic constructs that reside solely inside the heads of individual observers, or can observers reach agreement in applying trait terms?

2. Discriminativeness versus generality. If observers can agree with one another in ascribing traits to targets, is it simply because they apply a nondiscriminative "one size fits all" approach?

3. Behavior versus labeling. If observers can agree with one another, and can also differentiate between who is low or high on a given trait, does this occur because they really observe behavior? Or do they merely provide their judgments based on superficial stereotypes, targets' self-presentations, or other socially assigned labels?

4. Internal versus external locus of causal explanation. If observers can agree with one another and can distinguish individual differences on the basis of *actual behavior* of the people they are observing, are the causes of these consistencies located within each person or within his or her situation and role?

Each of these issues depends on the resolution of those earlier in the list. For instance, if observers cannot agree with one another about who has which traits, there is no point in going on to debate whether traits have a behavioral basis. Ultimately, assumptions about traits must pass the tests of consensus, discriminativeness, behavioral foundation, and internality. We will discuss seven hypotheses that assume that traits fail one or more of these tests. In Table 1, we list the hypotheses in terms of the four hierarchical issues just discussed. As can be seen, the hypotheses can be arranged more or less in order of their pessimism regarding the existence of (consensually verifiable, discriminative, internal) trait-like consistencies.[1]

We will consider each hypothesis in its purest form and, for the moment, disregard the various qualifications that have sometimes been attached to each. Placing each hypothesis in bold relief allows us to assess it most clearly, and philosophers of science tell us that we learn most when hypotheses are stated in such a way as to allow disproof (e.g., Platt, 1964; Popper, 1959). Moreover, each of these hypotheses has, at some time, actually been stated in its bold form. In 1968, for instance, one social psychologist argued that

Table 1
Hierarchy of Hypotheses From the Person–Situation Controversy, Arranged From Most to Least Pessimistic

Critical assumptions	Hypotheses
Solipsism over consensus	1. Personality is in the eye of the beholder.
Consensus without discrimination	2. Agreement between raters is an artifact of the semantic structure of the language used to describe personality. 3. Agreement is an artifact of base-rate accuracy (rater's tendency to make similar guesses about what people in general are like).
Discriminative consensus without behavioral referents	4. Differential agreement is an artifact of the shared use of invalid stereotypes. 5. Observers are in cahoots with one another; that is, their agreement results from discussion rather than accurate observation.
Differential agreement about behavior without internal traits	6. Raters see targets only within a limited range of settings and mistake situational effects for traits. 7. Compared with situational pressures, cross-situational consistencies in behavior are too weak to be important.

> the prevalent view that the normal behavior of individuals tends toward consistency is misconceived [and the research evidence] . . . strongly suggests that consistency, either in thought or action, does not constitute the normal state of affairs. (Gergen, 1968, pp. 305–306)

In the same year, a behavioral psychologist stated that "I, for one, look forward to the day when personality

Portions of this article were completed while the first author was visiting at Temple University and at the University of Tulsa. The research of the second author is supported by National Institute of Mental Health Grant R01-MH42427.

The authors thank Art Beaman, Jeff Goldstein, David Kenny, Dan Montello, John Reich, Melanie Trost, and Steve West for helpful comments on an earlier draft.

Correspondence concerning this article should be addressed to Douglas T. Kenrick, Department of Psychology, Arizona State University, Tempe, AZ 85287, or to David C. Funder, Department of Psychology, University of Illinois, 603 East Daniel, Champaign, IL 61820.

[1] These hypotheses are logically hierarchical in the sense that hypotheses at a later level do not arise if the earlier ones are true in their bold form. There must be some consensus, some discrimination, and some internality for trait research to be meaningful. However, each of the hypotheses points to a possible source of methodological error that cannot be ignored once one has passed to the next level. Although we will conclude that each hypothesis can be ruled out in its radical form, each one makes a unique methodological contribution. We will discuss the methodological warnings raised by each hypothesis as we proceed.

theories are regarded as historical curiosities" (Farber, 1964, p. 37).

Such extreme pessimism was clearly unwarranted. The data available now, more than two decades later, argue strongly against all seven of the hypotheses in Table 1. However, it would be a mistake to presume, as some personologists seem to do, that the issues raised by the "situationists" were merely diversions from the true path that can now be safely disregarded. We have learned, in the course of the debate, about a number of sources of distortion in trait judgments. These not only are of interest in their own right but are useful to personality assessment professionals, whose main goal may be to eliminate as much clutter from their path as possible.

Hypothesis 1: Personality Is in the Eye of the Beholder

The first and most pessimistic hypothesis that must be considered is that our perceptions of personality traits in our friends, acquaintances, and selves might be largely or exclusively by-products of the limitations and flaws of human information processing. Although no personality researcher has ever advocated that personality exists solely in the head and not in the external world, social psychologists such as Gergen (1968) and behavioral analysts such as Farber (1964) have done so. Moreover, the issue lies in the logical path of any further inquiries into the origin of trait attributions.

Social psychologists have often emphasized how personality impressions can arise in the absence of supporting evidence in the real world:

> Unwitting evidence provided by countless personality psychologists shows how objectively low or nonexistent covariations (between personality and behavior) can be parlayed into massive perceived covariations through a priori theories and assumptions (Nisbett & Ross, 1980, p. 109)

> The personality theorists' (and the layperson's) conviction that there are strong cross-situational consistencies in behavior may be seen as merely another instance of theory-driven covariation assessments operating in the face of contrary evidence. (Nisbett & Ross, 1980, p. 112)

Research relevant to the "eye of the beholder" hypothesis has mainly consisted of (a) demonstrations of various "errors" in the way that people process social information, or (b) claims that different judges rating the same personality rarely agree with each other or with the person being rated.

The demonstrations of error (for reviews, see Nisbett & Ross, 1980; Ross, 1977) establish that information given to subjects in laboratory settings is frequently distorted. People tend to jump to conclusions, biasing their judgments and their memories on the basis of their "implicit personality theories" (Schneider, 1973) or "scripts" (Abelson, 1976; Schank & Abelson, 1977). Studies of these attributional errors clearly demonstrate that people have biased expectations and that they routinely go beyond the information they are given.

However, for two reasons such studies do not establish that personality resides solely in the eye of the beholder. First, some of the errors are more a product of the unusual experimental situation than of a fundamentally biased cognitive process (cf. Block, Weiss, & Thorne, 1979; Trope, Bassok, & Alon, 1984). More important, the existence of judgmental biases does not necessarily imply the existence of mistakes. The expectations and biases demonstrated in laboratory tasks are, in principle, liable to lead to correct judgments in the real world (Funder, 1987). Many demonstrations of this principle can be found in the field of visual perception, where a useful rule of thumb underlies every "optical illusion" (Gregory, 1971). The "Ponzo" or "railroad lines" illusion, for example, produces errors in the lab but correct judgments when applied to three-dimensional reality. In the field of social perception, even the "fundamental attribution error" will lead to correct judgments to the extent that real people actually are somewhat consistent in their behavior. In short, demonstrations of laboratory errors are not informative, one way or the other, as to whether the associated judgmental biases lead mostly to mistakes or correct judgments in real life (see also McArthur & Baron, 1983).

A different line of support for the "eye of the beholder" hypothesis has been the belief that people generally do not agree with each other in their judgments of the same personality. For example, Dornbusch, Hastorf, Richardson, Muzzy, and Vreeland (1965) found that the constructs children in a summer camp used to describe personality were more a function of the person doing the ratings than they were of the person being rated. Such studies do show that people have individually preferred constructs for thinking about others. But these judgmental idiosyncrasies must be interpreted in the light of frequent findings that (a) when raters and ratees get a chance to know one another, their ratings come to agree with each other more (Funder & Colvin, 1987; Norman & Goldberg, 1966), and (b) when common rating categories are imposed on raters, their judgments will show substantial agreement in orderings of individual targets (e.g., Amelang & Borkenau, 1986; Bem & Allen, 1974; Cheek, 1982; Funder, 1987, Funder & Dobroth, 1987; Kenrick & Braver, 1982; Koretzky, Kohn, & Jeger, 1978; McCrae, 1982; Mischel & Peake, 1982).

Table 2 demonstrates some fairly typical findings in the area. In each of these studies, adult targets rated their own personalities and were also rated by more than one person who knew them well (parents, spouses, housemates, or friends). Correlations represent agreement about the same person by different raters who filled out the scales independently. Studies on the left side of the table used single-item scales (Funder & Dobroth, 1987; Kenrick & Stringfield, 1980); Dantchik (1985) and Cheek (1982) used 5-item and 3-item scales, respectively; and the studies to the right used lengthier scales with better established psychometric properties. It is clear that the use of reliable rating scales leads to high agreement regarding a target's personality, but even single-item scales can produce consistently positive (and statistically significant) levels of agreement. In fact, when the Kenrick

Table 2
Interrater Correlations From Recent Trait Studies

Trait	Kenrick & Stringfield (1980) (n = 71)	Kenrick & Stringfield (1980) Obs[a] (n = 34)	Funder & Dobroth (1987) (n = 69)	Dantchik (1985) (n = 92)	Dantchik (1985) Obs[a] (n = 36)	Cheek (1982) 1/2/3[b] (n = 81)	Cheek (1982) Obs[a] (n = 40)	McCrae (1982) (n = 139)	Paunonen & Jackson (1985) (n = 90)	Mischel & Peake (1982) (n = 63)
Intellectance	.17	.04	.36	.40	.52			.50	.53	
Likeability	.35	.52	.41	.14	.14	.22/.33/.39	.36	.47	.57	
Self-control	.26	.26	.25	.19	.47	.27/.40/.47	.49	.48	.67	.52
Sociability	.40	.55	.34	.46	.53	.43/.53/.59	.64	.53	.74	
Adjustment	.23	.43	.23	.38	.40	.22/.25/.27	.46	.58	.48	
Dominance	.35	.41	.40	.58	.61			.52	.60	
M	.29 (.53)[c]	.37 (.67)	.34	.37 (.51)	.45 (.64)	.29/.38/.44	.50	.51	.59	.52

Note. The trait labels used here are based on Hogan's (1982) terminology, and we have used roughly equivalent scales from studies that did not use those exact terms (denoting the major "factors" usually found in trait rating studies).
[a] Data from subjects who rated their behaviors on a given dimension as publicly observable (Obs).
[b] Data based on 1, 2, and 3 judges, respectively.
[c] Figures in parentheses are corrected for attenuation.

and Stringfield and the Dantchik data are corrected for attenuation (using test–retest unreliability estimates), the data fall right into line with those obtained in studies using more reliable scales (McCrae, 1982; Mischel & Peake, 1982; Paunonen & Jackson, 1985).[2] In line with these findings, Paunonen (1984) systematically varied the number of items in rating scales and found that agreement between target ratings and peer ratings rose from the .20s to the .50s as more items were added (and the rating scales thus became more reliable).

A consideration of this first hypothesis has taught us something about when the eyes of different beholders will behold different characteristics in the persons at whom they are looking. For instance, when rating strangers, observers will be quite happy to make attributions about what the strangers are like but will show little consensus (Funder & Colvin, 1987; Monson, Keel, Stephens, & Genung, 1982; Passini & Norman, 1966). So, although strangers' ratings provide an excellent domain for the study of bias (Fiske & Taylor, 1984), it is probably futile to expect them to manifest much validity. However, when observers are well acquainted with the person they are judging, they nevertheless do manage to see something on which they can agree. The findings of consensus (such as those in Table 2) are sufficient to rule out the radical hypothesis that personality resides solely in the eye of the beholder.

Recently, Kenny and La Voie (1984) showed how sophisticated statistical analyses can begin to separate the variance in ratings that is due to the target from that due to the rater's idiosyncratic biases. To separate these components, it is necessary to have multiple ratings of each target and multiple ratings (of different people) by each rater. In a preliminary analysis of several studies that met those criteria, Kenny and La Voie estimated that 41% of the variance was due to the person being rated and 17% was due to unique biases of the rater (the remaining variance was due to error or to unique interactions between target and rater). Their findings add to those already discussed in suggesting that there is probably less variance in the idiosyncratic eye of the beholder than there is in the consensual eye of the community. However, such consensus does not itself establish the accuracy of these ratings. People could agree with each other but still be wrong. Such interjudge agreement might be an artifact, and several possible candidates will now be considered.

Hypothesis 2: Agreement Is Due to Semantic Generalization

The first hypothesis, in its radical form, considered traits to be idiosyncratic constructions of the individual perceiver. The second hypothesis concedes that there is consensus in the use of trait terms but views that agreement as due simply to shared delusions based on common linguistic usage. According to the semantic generalization hypothesis, as soon as one judgment about another person is made, many other judgments follow based on nothing more than implicit expectations about which words "go together." Anyone judged as "friendly" may also be judged as "empathic," "altruistic," and "sincere" because the concepts are semantically linked, even though the component behaviors themselves may not be so linked. For instance, "helping others in distress," and "contributing to charities" (behavioral components of "altruism") may not be correlated with "smiling a lot" and "talking to strangers" (behavioral components of "friendliness"),

[2] Reliable ratings can also be obtained through the use of multiple observers (as Cheek's data demonstrate), and agreement also seems to be enhanced when the relevant criteria are publicly observable dimensions, as shown in the studies by Cheek (1982), Dantchik (1985), Funder and Colvin (1987), Funder and Dobroth (1987) and Kenrick and Stringfield (1980). We will return to these issues later, but for now the main point is that substantial interrater agreement in judgments of personality is not only possible but regularly found.

but judges who see evidence of "smiling a lot" might still infer "altruism," at least sometimes incorrectly. Shweder (1975) argued that shared preconceptions about "what goes with what" affected judgments so pervasively as to raise the question "How relevant is an individual differences theory of personality?" (See also D'Andrade, 1974.) Bourne (1977) went even further, suggesting that trait ratings might not reflect "anything more than raters' conceptual expectancies about which traits go together" (p. 863).

In the original studies, such effects seemed partly mediated by memory processes. When intercorrelations among immediate ratings of low-level behaviors were compared with intercorrelations among "memory ratings" and "semantic structure ratings," the memory and semantic structure ratings correlated more with one another than either one of these did with so-called "actual" behavior (D'Andrade, 1974; for evidence questioning the validity of these immediate behavior ratings, see Romer & Revelle, 1984; Rowe, 1982).

It is crucial to realize, as Block, Weiss, and Thorne (1979) pointed out, that semantic generalization cannot explain how different judges agree on attributing a *single* trait to a target person (as research such as that in Table 2 shows they do). To take a well-known example, the Passini and Norman (1966) study has been cited as evidence that trait ratings are based on "nothing more" than semantic similarity judgments. Indeed, Passini and Norman's data yielded a similar factor structure for ratings of friends and for ratings of strangers (who had been observed only briefly). Because the strangers had very little time to observe one another, it is clear that an implicit personality theory guided their judgments. However, this issue of the relationships between trait words is completely orthogonal to the question of accuracy in application of any one of those words. Passini and Norman's subjects not only reached significant agreement about which trait applied to which person but they also agreed more about friends' ratings than about strangers' (see also Funder & Colvin, 1987; Norman & Goldberg, 1966).

In light of such arguments, Shweder and D'Andrade (1979) seem to have reversed their earlier claim that semantic generalization negates the importance of judgments of individual differences. Although semantic structure might tell us to expect "friendly" to go with "altruistic" and not with "aggressive," it does not tell us whether we should apply the term more strongly to Walter or Seymour or Daryl. We must seek further for an adequate explanation of findings like those in Table 2.

Hypothesis 3: Agreement Is Due to Base-Rate Accuracy

According to this hypothesis, interrater agreement is an artifact of the highly stable base rates that many traits have in the population at large. For example, the trait "needs to be with other people" characterizes most of us, whereas "has murderous tendencies" characterizes few. If one is trying to describe someone one does not know, therefore, one can achieve a certain degree of "accuracy" just by rating the first trait higher than the second. The base-rate hypothesis, like the semantic structure hypothesis, allows for consensus between observers but regards their judgments as indiscriminate. "Accuracy" of this sort might reflect knowledge about what people in general are like, what Cronbach (1955) called "stereotype accuracy," but does not necessarily reflect any knowledge specific to the person being described.[3]

The base-rate accuracy problem helps us understand phenomena such as the "Barnum effect" (Ulrich, Stachnik, & Stainton, 1963), reflected in widespread acceptance of generalized descriptions such as, "You have a strong need for other people to like you and for them to admire you." Questions of when and for whom base-rate accuracy becomes an issue are interesting ones. For example, a recent study by Miller, McFarland, and Turnbull (1985) found that Barnum statements are more likely to be accepted by subjects when the statements refer to attributes that are publicly observable and flattering. However, to argue that base-rate accuracy is a basis for doubting whether we "can . . . describe an individual's personality" (Bourne, 1977) takes things too far. The base-rate accuracy hypothesis, like the semantic similarity hypothesis, can explain how judges reach consensus but not how they distinguish *between* the targets they judge. To take a simple case, imagine that a group of sorority sisters rates one another on a dichotomous item (as either "friendly" or "unfriendly"). If "friendly" is chosen over "unfriendly" 9 out of 10 times, there could be a very high percentage of "agreement," in terms of overlapping judgments, even if there were absolutely no agreement about who the 10th, unfriendly person is. But if there is truly no agreement about individual targets, correlations calculated between judges will show no relationship at all. So base-rate accuracy cannot explain the results of interrater studies such as those in Table 2 either (cf. Funder, 1980a; Funder & Colvin, 1987; Funder & Dobroth, 1987).

Summarizing thus far, we may say that whatever role solipsism and glittering generality play as noise in personality assessment, a signal of consensus and discrimination comes through. Can that signal be explained without acceding to the existence of trait-like consistencies in behavior? The answer is still yes, and in at least three ways.

Hypothesis 4: Agreement Is Due to Stereotypes Based on Obvious (but Erroneous) Cues

None of the arguments considered so far can account for interjudge agreement about the differences between people. One hypothesis that does is this: Perhaps agreement about peers is due to shared (but incorrect) stereotypes based on one or another readily accessible (e.g., physical) cues. Many such stereotypes come to mind: physical types (athlete, fat person, dumb blonde), racial and ethnic ste-

[3] The problem of "stereotype accuracy" was a large part of the reason why research on individual differences in accuracy in person perception largely died out following Cronbach's (1955) critique. The usual criterion for accuracy had been interprofile agreement scores that turned out to be centrally influenced by stereotype accuracy and other components.

reotypes, and so forth. Judges might share cultural stereotypes and so "agree" about burly, obese, or blond targets regardless of whether there were any corresponding consistencies in the targets' behavior.

Note that this hypothesis is very different from the sort of "stereotype accuracy" discussed under Hypothesis 3. That hypothesis referred to the possibility of indiscriminate responding based on raters' common preconceptions about what *everybody* is like. Hypothesis 4 refers to consensual agreement about traits that are *differentially* assigned to others. None of the first three hypotheses requires the observer to really "observe" anything distinctive about the person he or she is describing. This hypothesis, however, does require that the observer at least take a look at the target person—but assumes that the observer hardly looks much further than the end of his or her nose, just enough to assign the target person to a general category.

Such categorical stereotypes undoubtedly exist, but this does not mean we cannot become more accurate after getting to know someone beyond their "surface" categorization. Raters will try to make "reasonable" (i.e., stereotypic) guesses in the absence of real behavioral information. But as we mentioned earlier, their ratings increasingly converge as they actually observe the person's behavior (e.g., Funder & Colvin, 1987; Monson, Tanke, & Lund, 1980; Moskowitz & Schwarz, 1982; Norman & Goldberg, 1966; Passini & Norman, 1966).

The data that are most difficult for the stereotype hypothesis to explain are relationships between judgments and independent, objective behavioral measurements. For example, parents and teachers can provide general personality descriptions of children that not only agree with each other but also predict the children's "delay of gratification" behavior, measured in minutes and seconds, in a lab situation that none of the raters have ever seen (Bem & Funder, 1978; Funder, Block, & Block, 1983; Mischel, 1984). Other examples include Funder's studies of personality correlates of attributional style (1980b), attitude change (1982), and social acuity (Funder & Harris, 1986b); Gormly and Edelberg's (1974) work on aggression; Moskowitz and Schwarz's (1982) work on dominance; and Alker and Owen's (1977) research on reactions to stressful events. This sort of predictive capability must arise from something beyond the use of invalid stereotypes.

Although the existence of stereotypes does not negate the existence of traits, it is useful to consider how stereotypes and personality traits interact. For example, physical attractiveness may actually lead one to become more friendly, via self-fulfilling prophecies (Goldman & Lewis, 1977; Snyder, Tanke, & Berscheid, 1977). Likewise, burly males really are more aggressive (Glueck & Glueck, 1956), probably because aggressiveness has a higher payoff for a muscular youth than it does for a skinny or flabby one.

In sum, although stereotypes may be informative about the genesis of some traits, and may account for judgments of strangers, the findings that observers agree more with one another after they have gotten to know the target and the correlations between ratings and independent assessments of behavior rule out the possibility that interrater agreement is due solely to the use of shared stereotypes based on superficial cues.

Hypothesis 5: Agreement Is Due to Discussion Between Observers

We just considered evidence that observers agree with each other better when they know the target person well. Is this because acquaintances have had more time to observe the relevant behaviors and hence are more truly accurate than strangers? Perhaps not. It could be argued that observers ignore the truly relevant nonverbal behaviors of a target person but are attentive to the target's verbalizations about himself or herself and come to regard the target as the target does for that reason (cf. Funder, 1980a; Funder & Colvin, 1987). Alternatively, observers might get together and discuss the target (McClelland, 1972), agree on his or her reputation, and then inform the target about how to regard himself or herself (as in the classical "looking glass self" formulations of C. H. Cooley, 1902).

The research cited earlier, showing how ratings of personality traits can predict behavior in unique settings, strongly suggests that such explicit "negotiation" is not all that underlies interjudge agreement. Moreover, several researchers have found that agreement between parents "back home" and peers at college is about as good as that among peers or among parents (Bem & Allen, 1974; Kenrick & Stringfield, 1980). Likewise, Koretzky et al. (1978) found respectable agreement between judges from different settings. In that study, the various settings were all within the same (mental) institution, but the Kenrick and Stringfield (1980) study was conducted in an isolated college town in Montana and used parents who often lived several hundred miles away from campus and were unlikely to have met the peers (whose home towns may have been hundreds of miles in the opposite direction), much less to have had intimate discussions with them about their children's traits.

Findings of higher agreement on traits that relate to observable behaviors (such as "friendliness" as opposed to "emotionality") are also relevant here. Kenrick and Stringfield (1980) found that "observable" traits are reported with better agreement than "unobservable" ones. This tendency held both across traits (with some traits—like "shyness"—being nomothetically rated as observable) as well as within traits (with some subjects rating their "emotionality" as more observable than others did). Related findings are reported by Amelang and Borkenau (1986), Cheek (1982), Funder and Colvin (1987), Funder and Dobroth (1987), and McCrae (1982) and in two unpublished studies, one by Dantchik (1985) and one by McCall, Linder, West, and Kenrick (1985). If judges simply manufacture a reputation for a subject, it seems that it would be just as easy to agree about terms relating to emotionality as it would be to agree about terms relating to extraversion. Higher agreement about publicly observ-

able traits thus suggests that behavior is in fact being observed and accurately reported.

A tenacious adherent could still rescue this hypothesis by adding one more assumption. Perhaps we talk more about the so-called observable traits like extraversion than about "unobservable" traits. However, other findings further undermine the "discussion" hypothesis. Several studies have shown that when subjects' self-reports contradict their nonverbal behaviors, observers pay more attention to what is done than to what is said (Amabile & Kabat, 1982; Bryan & Walbek, 1970). In the Amabile and Kabat study, subjects viewed a target who described herself as either "introverted" or "extraverted", and they also watched her behave in a way that was either consistent with, or inconsistent with, her self-description. Observers' subsequent judgments were much more strongly influenced by her actual behaviors than by the way she had described herself. It seems, then, that observers give more credence to trait-relevant behaviors than to self-descriptions.

Summarizing our arguments thus far, there is good evidence that trait ratings are more than solipsistic fantasies. Observers can agree in their trait ratings and can use them differently for different people. For those we know well, at least, trait ratings involve more than just stereotypes based on easily observable categories, and they are based more on behavioral observation than on unfounded gossip. Are we therefore now compelled to allow some veracity to the trait construct? Alas, the answer is still no, not necessarily.

Hypothesis 6: Agreement Is Due to Seeing Others in the Same Setting

It is possible to allow for consensus and discrimination in the use of trait terms, and even to allow that observers are really and truly observing behavioral consistencies, without allowing that those behavioral consistencies stem from factors that are "internal" to the target person. As William James (1890), noted,

> Many a youth who is demure enough before his parents and teachers, swears and swaggers like a pirate among his "tough" young friends. We do not show ourselves to our children as to our club-companions, to our customers as to the laborers we employ, to our masters and employers as to our intimate friends. (p. 294)

Fellow club-companions may all agree that a particular merchant is consistently rather "wild," whereas his customers agree that he is quite "conventional." Because club-companions and customers live in "separate worlds," their different mutual delusions about the merchant's traits can be maintained. If behavior is mostly due to the situation, then the people who inhabit a given situation with a target will agree about that person's behavioral attributes, even if they are not actually general attributes of the individual's personality.

A good deal of the evidence we have already discussed poses difficulties for this hypothesis as a final explanation of rater agreement. Much of the research that uses trait ratings is based on studies of students who are rated by fellow fraternity members or college roomates (e.g., Bem & Allen, 1974; Cheek, 1982; Funder, 1980a; Funder & Dobroth, 1987; Kenrick & Stringfield, 1980). These individuals see each other across many settings, yet agree well. Recall also that studies such as those done by Bem and Allen (1972) and Kenrick and Stringfield (1980) found agreement across peer and parent groups—who see the targets in very different situations. In the Kenrick and Stringfield (1980) study, for instance, peers knew the target as a college student (and perhaps fellow beer-drinker), whereas parents knew the target as a child (and perhaps a ranch-hand). Restriction of range of environmental experience could even constrain correlations. For example, perhaps the college dorm is a setting that constrains one to be "friendly." If so, it will be a difficult and subtle task for raters who know two targets only in that setting to agree about which one is the more "dispositionally" friendly.[4]

Finally, a good deal of the research just discussed shows how personality ratings made by parents, teachers, and friends often correlate well with behavior measured in settings that are very different from the contexts from which their judgments were derived. From observing their children at home, for example, parents can provide personality descriptions that predict behavior measured in a unique experimental setting (Bem & Funder, 1978)—even when a dozen or more years separate the personality judgments from the behavior (Mischel, 1984). Such predictability has to be based on the parents' detection of true "cross-situational consistency."

Although the "situational" hypothesis is often viewed as an alternative to the trait position, they need not be at odds with one another. Researchers have begun to uncover useful information about how persons and situations "interact" (e.g., Bem & Funder, 1978; Kenrick & Dantchik, 1983; Magnusson & Endler, 1977; Snyder & Ickes, 1985):

1. Traits influence behavior only in relevant situations (Allport, 1966; Bem & Funder, 1978). Anxiety, for example, shows up only in situations that the person finds threatening.

2. A person's traits can change a situation (Rausch, 1977). For instance, an aggressive child can bring out the hostility in a previously peaceful playground.

3. People with different traits will choose different settings (Snyder & Ickes, 1985). Highly sex-typed males,

[4] Psychological tests can themselves be regarded as situations. Along these lines, Campbell and Fiske (1959) discussed the contribution of method variance to some of the high intercorrelations that are found, and Mischel (1968) noted that this can explain a number of high correlations in the personality literature. It is important to note that it is one sort of thing to say that "method variance" can produce a high correlation when Individual A rates himself on two different measures of Trait X, or when a clinician rates that person on two different measures, and quite another to talk about "method variance" when two different observers rate Individual A on the same measure. The former can be easily explained as "spurious" due to self-presentational consistency only, but the latter requires objective agreement about something about A and his or her behavior (assuming, of course, that the judges did not confer over their ratings).

for example, seek out sexually stimulating situations; highly sex-typed females avoid them (Kenrick, Stringfield, Wagenhals, Dahl, & Ransdell, 1980).

4. Traits can change with chronic exposure to certain situations. For instance, Newcomb's students became less conservative during their Bennington college experience and stayed that way for decades (Newcomb, Koenig, Flacks, & Warwick, 1967).

5. Traits are more easily expressed in some situations than others. They have more influence when situations are low in constraint—for example, a picnic as opposed to a funeral (Monson et al., 1982; Price & Bouffard, 1974; Schutte, Kenrick, & Sadalla, 1985). Traits are also more likely to be influential in settings that are highly prototypical or exemplary (Schutte et al., 1985). For instance, the postinterview cocktail party for an academic job applicant is more difficult to categorize than the in-office interview or the office Christmas party and would probably allow for the operation of greater individual differences. Note that laboratory situations, where psychologists often look for evidence of individual differences, will constrain the operation of traits precisely because they are rigidly controlled, are imposed arbitrarily on subjects, and are usually not reactive to anything the subject does (Monson & Snyder, 1977; Wachtel, 1973).

The data we have discussed thus far require us to concede that some degree of consensus, discrimination, and internality exist in the trait domain. Is it time, therefore, to give the store back to the "trait" position? Even with the distance we have come, the answer is still no. It is possible to argue that although some true cross-situational consistencies in behavior may exist, they are too small to worry about.

Hypothesis 7: The Relationships Between Traits and Behavior Are "Too Small" To Be Important

Just how small is "too small"? Mischel's (1968) review concluded that correlations between trait scores and behaviors and between different behaviors are seldom larger than about .30. This conclusion hit the field of personality with devastating force because of two separable assumptions: (a) The coefficient .30 is not simply an artifact of poorly developed research tools but is the true upper limit for the predictability of behavior from personality, and (b) this upper limit is a small upper limit. Acceptance of both of these assumptions was necessary for Mischel's critique to have had a major impact, and many initially did accept them.

Several personologists (e.g., Block, 1977; Hogan, DeSoto, & Solano, 1977) have challenged the first assumption, arguing that Mischel's review did not give a fair hearing to the better studies in the personality literature. More than the several studies cited in earlier sections of this article have used direct behavioral observations and found larger correlations with behavior (see also Block, Buss, Block, & Gjerde, 1981; Block, von der Lippe, & Block, 1973; McGowen & Gormly, 1976; Moskowitz, 1982). Epstein (1979, 1983) reported that such correlations can be especially high when aggregates of behavior rather than single instances are used.

Indeed, in everyday life, what we usually wish to predict on the basis of our personality judgments are not single acts, but aggregate trends: Will this person make an agreeable friend, a reliable employee, an affectionate spouse? Given such broad criteria, the Spearman-Brown formula shows how even "small" single-act correlations compound into extremely high predictive validities. For example, Mischel and Peake (1982) found that interitem correlations between behavior measures are relatively low (.14 to .21) for single, unaggregated observations but that coefficient alpha for their total behavioral aggregate is .74. That is, a similar aggregate of behaviors would correlate .74 with that one. Along the same lines, Epstein and O'Brien (1985) reanalyzed several classical studies in the field of personality. In all of these studies behavior was situation specific at the single-item level (in line with Mischel's point) but cross-situationally general at the level of behavioral aggregates. Protagonists on both sides of the controversy now seem ready to allow that the ".30 ceiling" applies only to behavior in unaggregated form (Epstein, 1983; Mischel, 1983).[5]

Even if one were to allow that it is difficult to surpass correlations of .30 to .40 (e.g., in the case of unaggregated measures), it may be a mistake to assume that such correlations are "small." In fact, correlations in this range characterize the strength of some of the most interesting and important situational effects found by experimental social psychology (Funder & Ozer, 1983; Sarason, Smith, & Diener, 1975) and even some of Mischel's own work on situational determinants of delay of gratification behavior (Funder & Harris, 1986a). These observations echo Hogan et al.'s (1977) warning that a correlation of .30 does not necessarily mean that the "remaining 91% of the variance" can be assigned, by subtraction, to the situation.

Moreover, a correlation of .30 may not be as small as many psychologists seem to believe. Common practice, as exemplified in the above warning, is to square such a correlation and report that it "accounts for 9% of the variance." However, Ozer (1985) claimed that, contrary to common belief and practice, the unsquared correlation coefficient is directly interpretable as the percentage of the variance accounted for: For example, $r = .30$ accounts for 30%, not 9%, of the relevant variance. Another way of clarifying the size of an effect in this range is Rosenthal and Rubin's (1979, 1982) binomial effect size display, which reveals that a predictor that correlates .30 with a dichotomous criterion will yield correct discriminations 65% of the time. Abelson (1985) made the point in a vivid way with an application of the "percentage of variance" approach to batting performances in major league

[5] Mischel (1985) recently reported behavioral consistency coefficients for aggressiveness in children that are consistently above .50 without aggregation across different situations. Mischel's position is not that consistency never exists. Instead, he now argues that adequate predictive validity can be obtained if we take careful account of situational factors that will elicit differential consistencies.

baseball players. Noting that most are in the .200s to .300s, he calculates that the percentage of variance explained in a single batting performance is less than 1%. Yet, with aggregation over seasons, these miniscule differences compound to result in discriminations important enough to determine hundreds of thousands of dollars in salary differentials. Thus, the .30 statistic that had such a devastating effect on the enterprise of personality assessment may have been badly misunderstood.

The hypothesis that personality coefficients are "too small" has been quite useful in elucidating some important limitations on what can be measured and how it should be measured. Minute and unaggregated behavioral indexes, no matter what their face validity, are not necessarily good criterion measures (Golding, 1978; West, 1983). They may be full of various sorts of error, lack temporal stability, or measure something other than what they seem to measure (Bem & Funder, 1978; Moskowitz & Schwarz, 1982; Romer & Revelle, 1984). Even if it is true, as Fiske (1979) pointed out, that judges can agree quite well about the occurrence of a given facial twitch, the twitch may be meaningless unless its context is understood (Block et al., 1979; Dahlstrom, 1972; Hogan, DeSoto & Solano, 1977). These problems may account for the repeated finding that when objective behavioral measures are compared with observers' ratings, the results do not support the superiority of behavioral measures (e.g., Eaton & Enns, 1986; Moskowitz & Schwarz, 1982).

What Have We Learned?

As with most controversies, the truth finally appears to lie not in the vivid black or white of either extreme, but somewhere in the less striking gray area. It would be a mistake, however, to claim that the interchange served only to bring out a number of "straw man" positions that no one ever took seriously anyway, that the repetitive cycle of argument and reply produced no more than fatigue and déjà vu, or that we are no closer to understanding personality traits than we were two decades ago. Radical versions of each of these hypotheses were suggested, not just for rhetorical purposes, and were passed uncritically onward to a generation of students in psychology courses. We were trained as experimental social psychologists during the heat of the debate, and the shade of gray we see now seemed much closer to a gloomy black back then. Indeed, for a time, and in some places, it was not unusual for the very idea of personality traits to be dismissed out of hand and even ridiculed.

On the other hand, one of us also underwent clinical training during that era and came across a viewpoint much closer to the "pure trait" position than is remotely tenable on the basis of the data available now. Ten years ago, there were, and probably still are (Mischel, 1983; Wade & Baker, 1977), clinical professionals overconfidently making grand predictions from minute samples of behavior of highly questionable reliability and validity. We can eliminate the radical forms of each of the seven critical hypotheses, but that does not imply that the so-called "pure trait" position has regained the day. Systematic sources of judgmental bias, systematic effects of situations, and systematic interactions between persons and situations must be explicitly dealt with before we can predict from trait measures.

So although there may be enough signal amidst the noise in this research area to make it worthwhile to turn on the radio, the device must still be carefully tuned. Instead of simply viewing each of the seven critical hypotheses as being resolved in favor of the trait position, it is better to view each as a clue about one ever-present source of noise to be tuned out. Kenny and La Voie (1984) showed how factors such as idiosyncratic rater bias (the problem of Hypothesis 1) can even, under the proper circumstances, be turned to statistical advantage in estimating a person's "true" trait score.

Other practical lessons have emerged from this controversy. The research now indicates quite clearly that anyone who seeks predictive validity from trait ratings will do better to use (a) raters who are thoroughly familiar with the person being rated; (b) multiple behavioral observations; (c) multiple observers; (d) dimensions that are publicly observable; and (e) behaviors that are relevant to the dimension in question.

On the other hand, one should *not* expect great accuracy when predicting (a) behavior in "powerful" and clearly normatively scripted situations from trait ratings and (b) a single behavioral instance from another single behavioral instance.

Those who would respond to this list by claiming that they "knew it all along" may or may not be guilty of hindsight bias (Fischoff, 1975). But they should at least acknowledge that many of us did not know these principles all along and needed the light generated by controversy to open our eyes. For instance, the apparently "obvious" insight that we should not rely on ratings made by strangers can help us understand why some of the data on clinical assessment (e.g., Goldberg & Werts, 1966; Golden, 1964; Soskin, 1959) have been so disappointing, and the awareness that traits will not show up in overpowering situations has led to a dramatic reassessment of failures to find "consistency" in brief laboratory observations. Likewise, if these issues and that of the unreliability of single behavioral instances were so obvious, one is left to wonder why the field responded so strongly to Mischel's (1968) critique. "Déjà vu" may be an accurate description of our current situation after all, because the term actually refers to the *illusion* that one has previously experienced something that is really new.

One side effect of the person–situation debate has been an intensification of the antagonism between personality and social psychology. Social psychologists have historically focused on situational determinants of behavior and were therefore quite willing to join with behavioral clinicians in the situationist attack on personality (Hogan & Emler, 1978; Kenrick & Dantchik, 1983). Personologists share a very different set of assumptions, and the two subdisciplines have sometimes seemed intent on defining each other out of existence (Kenrick, 1986). To continue such separation between the two fields would

be a mistake. Many exciting developments are beginning to emerge at the interface of social and personality psychology. For instance, research that combines personality with biology suggests a vast array of questions about the connection between personality traits and social interaction (Kenrick, 1987; Kenrick & Trost, 1987; Sadalla, Kenrick, & Vershure, 1987). And research on the accuracy of interpersonal judgment draws equally on both personality and social psychology (Funder, 1987; Funder & Colvin, 1987; Funder & Dobroth, 1987).

Houts, Cook, and Shadish (1986) made a strong case that science best progresses through multiple and mutually critical attempts to understand the same problem. When camps with strongly opposing sets of biases manage to come to some level of agreement, we may be more confident of the validity of the conclusions that are agreed upon. Viewed in this light, the controversy stimulated by the situationist attack on personality may be seen more as a life-giving transfusion than as a needless bloodletting.

REFERENCES

Abelson, R. P. (1976). A script theory of understanding, attitude, and behavior. In J. Carroll & J. Payne (Eds.), *Cognition and social behavior* (pp. 33–45). Hillsdale, NJ: Erlbaum.

Abelson, R. P. (1985). A variance explanation paradox: When a little is a lot. *Psychological Bulletin, 97,* 129–133.

Alker, H. A. (1972). Is personality situationally specific or intrapsychically consistent? *Journal of Personality, 40,* 1–16.

Alker, H. A., & Owen, D. W. (1977). Biographical, trait, and behavioral-sampling predictions of performance in a stressful life setting. *Journal of Personality and Social Psychology, 35,* 717–723.

Allport, G. W. (1931). What is a trait of personality? *Journal of Abnormal and Social Psychology, 25,* 368–372.

Allport, G. W. (1966). Traits revisited. *American Psychologist, 21,* 1–10.

Alston, W. P. (1975). Traits, consistency, and conceptual alternatives for personality theory. *Journal for the Theory of Social Behavior, 5,* 17–48.

Amabile, T. M., & Kabat, L. G. (1982). When self-description contradicts behavior: Actions do speak louder than words. *Social Cognition, 1,* 311–335.

Amelang, M., & Borkenau, P. (1986). The trait concept: Current theoretical considerations, empirical facts, and implications for personality inventory construction. In A. Angleitner & J. S. Wiggins (Eds.), *Personality assessment via questionnaire* (pp. 7–24). Berlin: Springer-Verlag.

Argyle, M., & Little, B. R. (1972). Do personality traits apply to social behavior? *Journal for the Theory of Social Behavior, 2,* 1–35.

Bem, D. J. (1972). Constructing cross-situational consistencies in behavior: Some thoughts on Alker's critique of Mischel. *Journal of Personality, 40,* 17–26.

Bem, D. J. (1983). Further *déjà vu* in the search for cross situational consistency: A reply to Mischel and Peake. *Psychological Review, 90,* 390–393.

Bem, D. J., & Allen, A. (1974). On predicting some of the people some of the time: The search for cross-situational consistencies in behavior. *Psychological Review, 81,* 506–520.

Bem, D. J., & Funder, D. C. (1978). Predicting more of the people more of the time: Assessing the personality of situations. *Psychological Review, 85,* 485–501.

Block, J. (1968). Some reasons for the apparent inconsistency of personality. *Psychological Bulletin, 70,* 210–212.

Block, J. (1977). Advancing the science of personality: Paradigmatic shift or improving the quality of research? In D. Magnusson & N. S. Endler (Eds.), *Personality at the crossroads: Current issues in interactional psychology* (pp. 37–63). Hillsdale, NJ: Erlbaum.

Block, J., Buss, D. M., Block, J. M., & Gjerde, P. F. (1981). The cognitive style of breadth of categorization: The longitudinal consistency of personality correlates. *Journal of Personality and Social Psychology, 40,* 770–779.

Block, J., von der Lippe, A., & Block, J. H. (1973). Sex-role and socialization patterns: Some personality concomitants and environmental antecedents. *Journal of Consulting and Clinical Psychology, 41,* 321–341.

Block, J., Weiss, D. S., & Thorne, A. (1979). How relevant is a semantic similarity interpretation of personality ratings? *Journal of Personality and Social Psychology, 37,* 1055–1074.

Bourne, E. (1977). Can we describe an individual's personality? Agreement on stereotype versus individual attributes. *Journal of Personality and Social Psychology, 35,* 863–872.

Bowers, K. S. (1973). Situationism in psychology: An analysis and critique. *Psychological Review, 80,* 307–336.

Bryan, J., & Walbek, N. (1970). Impact of words and deeds concerning altruism upon children. *Child Development, 41,* 747–757.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait–multimethod matrix. *Psychological Bulletin, 56,* 81–105.

Carlson, R. (1984). What's social about social psychology? Where's the person in personality research? *Journal of Personality and Social Psychology, 35,* 1055–1074.

Cheek, J. M. (1982). Aggregation, moderator variables, and the validity of personality tests: A peer-rating study. *Journal of Personality and Social Psychology, 43,* 1254–1269.

Cooley, C. H. (1902). *Human nature and the social order.* New York: Scribner's.

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

Dahlstrom, W. G. (1972). *Personality systematics and the problem of types.* Morristown, NJ: General Learning Press.

D'Andrade, R. G. (1974). Memory and the assessment of behavior. In H. M. Blalock (Ed.), *Measurement in the social sciences* (pp. 159–186). Chicago: Aldine-Atherton.

Dantchik, A. (1985). *Idiographic approaches to personality assessment.* Unpublished master's thesis, Arizona State University, Tempe.

Dornbusch, S. M., Hastorf, A. H., Richardson, S. A., Muzzy, R. E., & Vreeland, R. S. (1965). The perceiver and perceived: Their relative influence on categories of interpersonal perception. *Journal of Personality and Social Psychology, 1,* 434–440.

Eaton, W. D., & Enns, L. R. (1986). Sex differences in human activity level. *Psychological Bulletin, 100,* 19–28.

Endler, N. S. (1973). The person vs. situation: A pseudo issue? *Journal of Personality, 41,* 287–303.

Epstein, S. (1977). Traits are alive and well. In D. Magnusson & N. S. Endler (Eds.), *Personality at the crossroads: Current issues in interactional psychology* (pp. 83–98). Hillsdale, NJ: Erlbaum.

Epstein, S. (1979). The stability of behavior: I. On predicting most of the people much of the time. *Journal of Personality and Social Psychology, 37,* 1097–1126.

Epstein, S. (1980). The stability of behavior: II. Implications for psychological research. *American Psychologist, 35,* 790–806.

Epstein, S. (1983). The stability of confusion: A reply to Mischel and Peake. *Psychological Review, 90,* 390–393.

Epstein, S., & O'Brien, E. J. (1985). The person–situation debate in historical and current perspective. *Psychological Bulletin, 98,* 513–537.

Farber, I. E. (1964). A framework for the study of personality as a behavioral science. In P. Worchel & D. Bryne (Eds.), *Personality change* (pp. 3–37). New York: Wiley.

Fischoff, B. (1975). Hindsight does not equal foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance, 1,* 288–299.

Fiske, D. W. (1974). The limits for the conventional science of personality. *Journal of Personality, 42,* 1–11.

Fiske, D. W. (1979). Two worlds of psychological phenomena. *American Psychologist, 34,* 733–739.

Fiske, S., & Taylor, S. (1984). *Social cognition.* New York: Random House.

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

Funder, D. C. (1980b). The "trait" of ascribing traits: Individual differences in the tendency to trait ascription. *Journal of Research in Personality, 14,* 376–385.

Funder, D. C. (1982). On assessing social psychological theories through the study of individual differences: Template matching and forced compliance. *Journal of Personality and Social Psychology, 43,* 100–110.

Funder, D. C. (1983). Three issues in predicting more of the people: A reply to Mischel and Peake. *Psychological Review, 90,* 283–289.

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

Funder, D. C., Block, J., & Block, J. H. (1983). Delay of gratification: Some longitudinal personality correlates. *Journal of Personality and Social Psychology, 44,* 1198–1213.

Funder, D. C., & Colvin, C. R. (1987). *Friends and strangers: Acquaintanceship, agreement, and the accuracy of personality judgment.* Manuscript submitted for publication.

Funder, D. C., & Dobroth, J. M. (1987). Differences between traits: Properties associated with interjudge agreement. *Journal of Personality and Social Psychology, 52,* 409–418.

Funder, D. C., & Harris, M. J. (1986a). Experimental effects and person effects in delay of gratification. *American Psychologist, 41,* 476–477.

Funder, D. C., & Harris, M. J. (1986b). On the several facets of personality assessment: The case of social acuity. *Journal of Personality, 54,* 528–550.

Funder, D. C., & Ozer, D. J. (1983). Behavior as a function of the situation. *Journal of Personality and Social Psychology, 44,* 107–112.

Gergen, K. J. (1968). Personal consistency and the presentation of self. In C. Gordon & K. J. Gergen (Eds.), *The self in social interaction* (pp. 299–308). New York: Wiley.

Glueck, S., & Glueck, E. (1956). *Physique and delinquency.* New York: Harper & Row.

Goldberg, L. R., & Werts, C. E. (1966). The reliability of clinician's judgments: A multitrait–multimethod approach. *Journal of Consulting Psychology, 30,* 199–206.

Golden, M. (1964). Some effects of combining psychological tests on clinical inferences. *Journal of Consulting Psychology, 28,* 440–446.

Golding, S. L. (1978). Toward a more adequate theory of personality: Psychological organizing principles. In H. London (Ed.), *Personality: A new look at metatheories* (pp. 69–96). New York: Wiley.

Goldman, W., & Lewis, P. (1977). Beautiful is good: Evidence that the physically attractive are more socially skilled. *Journal of Experimental Social Psychology, 13,* 125–130.

Gormly, J., & Edelberg, W. (1974). Validity in personality trait attributions. *American Psychologist, 29,* 189–193.

Gregory, R. L. (1971). Visual illusions. In R. C. Atkinson (Ed.), *Contemporary psychology* (pp. 167–177). San Francisco: W. H. Freeman.

Hogan, R. (1982). A socioanalytic theory of personality. In R. A. Dienstbier & M. M. Page (Eds.), *Nebraska Symposium on Motivation* (Vol. 30, pp. 55–89). Lincoln: University of Nebraska Press.

Hogan, R., DeSoto, C. B., and Solano, C. (1977). Traits, tests, and personality research. *American Psychologist, 32,* 255–264.

Hogan, R. T., & Emler, N. P. (1978). The biases in contemporary social psychology. *Social Research, 45,* 478–534.

Houts, A. C., Cook, T. D., & Shadish, W. R. (1986). The person–situation debate: A critical multiplist perspective. *Journal of Personality, 54,* 52–105.

Hunt, J. McV. (1965). Traditional personality theory in the light of recent evidence. *American Scientist, 53,* 80–96.

Ichheisser, G. (1943). Misinterpretations of personality in everyday life and the psychologist's frame of reference. *Character and Personality, 12,* 145–160.

Jackson, D. N. (1983). Some preconditions for valid person perception. In M. P. Zanna, E. T. Higgins, & C. P. Herman (Eds.), *Consistency in social behavior: The Ontario Symposium* (pp. 251–279). Hillsdale, NJ: Erlbaum.

James, W. (1890). *Principles of psychology* (Vol. 1). London: Macmillan.

Kenny, D. A., & La Voie, L. (1984). The social relations model. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 18, pp. 141–182). Orlando, FL: Academic Press.

Kenrick, D. T. (1986). How strong is the case against contemporary social and personality psychology? A response to Carlson. *Journal of Personality and Social Psychology, 50,* 839–844.

Kenrick, D. T. (1987). Gender, genes, and the social environment. In P. C. Shaver & C. Hendrick (Eds.), *Review of personality and social psychology: Vol. 7. Sex and gender* (pp. 14–43). Beverly Hills, CA: Sage.

Kenrick, D. T., & Braver, S. L. (1982). Personality: Idiographic and nomothetic! A rejoinder. *Psychological Review, 89,* 182–186.

Kenrick, D. T., & Dantchik, A. (1983). Interactionism, idiographics, and the social psychological invasion of personality. *Journal of Personality, 51,* 286–307.

Kenrick, D. T., & Stringfield, D. O. (1980). Personality traits and the eye of the beholder: Crossing some traditional philosophical boundaries in the search for consistency in all of the people. *Psychological Review, 87,* 88–104.

Kenrick, D. T., Stringfield, D. O., Wagenhals, W. L., Dahl, R. H., & Ransdell, H. J. (1980). Sex differences, androgyny, and approach responses to erotica: A new variation on the old volunteer problem. *Journal of Personality and Social Psychology, 40,* 1039–1056.

Kenrick, D. T., & Trost, M. R. (1987). A biosocial theory of heterosexual relationships. In K. Kelley (Ed.), *Males, females, and sexuality: Theory and research* (pp. 59–100). Albany: State University of New York Press.

Koretzky, M. B., Kohn, M., & Jeger, A. M. (1978). Cross-situational consistency among problem adolescents: An application of the two-factor model. *Journal of Personality and Social Psychology, 36,* 1054–1059.

Magnusson D., & Endler, N. S. (Eds.). (1977). *Personality at the crossroads: Current issues in interactional psychology.* Hillsdale, NJ: Erlbaum.

McArthur, L. Z., & Baron, R. M. (1983). Toward an ecological theory of social perception. *Psychological Review, 90,* 215–235.

McCall, M., Linder, D. E., West, S. G., & Kenrick, D. T. (1985). *Some cautions on the template-matching approach to assessing person/environment interactions.* Unpublished manuscript, Arizona State University, Tempe.

McClelland, D. C. (1972). Opinions reflect opinions: So what else is new? *Journal of Consulting and Clinical Psychology, 38,* 325–326.

McCrae, R. R. (1982). Consensual validation of personality traits: Evidence from self-reports and ratings. *Journal of Personality and Social Psychology, 43.* 293–303.

McGowen, J., & Gormly, J. (1976). Validation of personality traits: A multicriteria approach. *Journal of Personality and Social Psychology, 34,* 791–795.

Miller, D. T., McFarland, C., & Turnbull, W. (1985). *Pluralistic ignorance: Its causes and consequences.* Paper presented at the annual meeting of the Eastern Psychological Association, Boston.

Mischel, W. (1968). *Personality and assessment.* New York: Wiley.

Mischel, W. (1983). Alternatives in the pursuit of the predictability and consistency of persons: Stable data that yield unstable interpretations. *Journal of Personality, 51,* 578–604.

Mischel, W. (1984). Convergences and challenges in the search for consistency. *American Psychologist, 39,* 351–364.

Mischel, W. (1985, October). *Diagnosticity of situations.* Paper presented at the meeting of the Society for Experimental Social Psychology, Evanston, IL.

Mischel, W., & Peake, P. K. (1982). Beyond *déjà vu* in the search for cross-situational consistency. *Psychological Review, 89,* 730–755.

Mischel, W., & Peake, P. K. (1983). Some facets of consistency: Replies to Epstein, Funder, and Bem. *Psychological Review, 90,* 394–402.

Monson, T. C., Keel, R., Stephens, D., & Genung, V. (1982). Trait attributions: Relative validity, covariation with behavior, and prospect of future interaction. *Journal of Personality and Social Psychology, 42,* 1014–1024.

Monson, T. C., Tanke, E. D., & Lund, J. (1980). Determinants of social perception in a naturalistic setting. *Journal of Research in Personality, 14,* 104–120.

Monson, T. C., & Snyder, M. (1977). Actors, observers, and the attribution

process: Toward a reconceptualization. *Journal of Experimental Social Psychology, 13,* 89–111.

Moskowitz, D. S. (1982). Coherence and cross-situational generality in personality: A new analysis of old problems. *Journal of Personality and Social Psychology, 43,* 754–768.

Moskowitz, D. S., & Schwarz, J. C. (1982). Validity comparison of behavior counts and ratings by knowledgeable informants. *Journal of Personality and Social Psychology, 42,* 518–528.

Newcomb, T. M., Koenig, K. E., Flacks, R., & Warwick, D. P. (1967). *Persistence and change: Bennington College and its students after twenty-five years.* New York: Wiley.

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

Norman, W. T., & Goldberg, L. R. (1966). Raters, ratees, and randomness in personality structure. *Journal of Personality and Social Psychology, 4,* 681–691.

Ozer, D. J. (1985). Correlation and the coefficient of determination. *Psychological Bulletin, 97,* 307–315.

Passini, F. T., & Norman, W. T. (1966). A universal conception of personality structure? *Journal of Personality and Social Psychology, 4,* 44–49.

Paunonen, S. V. (1984). Optimizing the validity of personality assessments: The importance of aggregation and item content. *Journal of Research in Personality, 18,* 411–431.

Paunonen, S. V., & Jackson, D. N. (1985). Idiographic measurement strategies for personality and prediction: Some unredeemed promissory notes. *Psychological Review, 92,* 486–511.

Platt, J. R. (1964). Strong inference. *Science, 146,* 347–353.

Popper, K. (1959). *The logic of scientific discovery.* New York: Basic Books.

Price, R. H., & Bouffard, D. L. (1974). Behavioral appropriateness and situational constraint as dimensions of social behavior. *Journal of Personality and Social Psychology, 30,* 579–586.

Rausch, M. L. (1977). Paradox, levels, and junctures in person-situation systems. In D. Magnusson & N. S. Endler (Eds.), *Personality at the crossroads* (pp. 287–304). Hillsdale, NJ: Erlbaum.

Romer, D., & Revelle, W. (1984). Personality traits: Fact or fiction? A critique of the Shweder and D'Andrade systematic distortion hypothesis. *Journal of Personality and Social Psychology, 47,* 1028–1042.

Rosenthal, R., & Rubin, D. B. (1979). A note on percent variance explained as a measure of the importance of effects. *Journal of Applied Social Psychology, 9,* 385–396.

Rosenthal, R., & Rubin, D. B. (1982). A simple, general purpose display of magnitude of experimental effect. *Journal of Educational Psychology, 74,* 166–169.

Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 10, pp. 174–221). New York: Academic Press.

Rowe, D. (1982). Monozygotic twins cross-correlations as a validation of personality structure: A test of the semantic bias hypothesis. *Journal of Personality and Social Psychology, 43,* 1075–1079.

Sadalla, E. K., Kenrick, D. T., & Vershure, B. (1987). Dominance and heterosexual attraction. *Journal of Personality and Social Psychology, 52,* 730–738.

Sarason, I. G., Smith, R. E., & Diener, E. (1975). Personality research: Components of variance attributable to the person and the situation. *Journal of Personality and Social Psychology, 32,* 199–204.

Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals, and understanding.* Hillsdale, NJ: Erlbaum.

Schneider, D. (1973). Implicit personality theory: A review. *Psychological Bulletin, 79,* 294–309.

Schutte, N. A., Kenrick, D. T., & Sadalla, E. K. (1985). The search for predictable settings: Situational prototypes, constraint, and behavioral variation. *Journal of Personality and Social Psychology, 49,* 121–128.

Shweder, R. A. (1975). How relevant is an individual-difference theory of personality? *Journal of Personality, 43,* 455–485.

Shweder, R. A., & D'Andrade, R. G. (1979). Accurate reflection or systematic distortion: A reply to Block, Weiss, and Thorne. *Journal of Personality and Social Psychology, 37,* 1075–1084.

Snyder, M., & Ickes, W. (1985). Personality and social behavior. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology* (3rd ed., Vol. 2, pp. 883–948). Reading, MA: Addison-Wesley.

Snyder, M., Tanke, E. D., & Berscheid, E. (1977). Social perception and interpersonal behavior: On the self-fulfilling nature of social stereotypes. *Journal of Personality and Social Psychology, 35,* 656–666.

Soskin, W. F. (1959). Influence of four types of data on diagnostic conceptualization in psychological testing. *Journal of Abnormal and Social Psychology, 58,* 69–78.

Trope, Y., Bassok, M., & Alon, E. (1984). The questions lay interviewers ask. *Journal of Personality, 52,* 90–106.

Ulrich, R. E., Stachnik, T. J., & Stainton, N. R. (1963). Student acceptance of generalized personality interpretations. *Psychological Reports, 13,* 831–834.

Wachtel, P. (1973). Psychodynamics, behavior therapy, and the implacable experimenter: An inquiry into the consistency of personality. *Journal of Abnormal Psychology, 82,* 324–334.

Wade, T. C., & Baker, T. B. (1977). Opinions and use of psychological tests: A survey of clinical psychologists. *American Psychologist, 32,* 874–882.

West, S. G. (1983). Personality and prediction: An introduction. *Journal of Personality, 51,* 275–285.

Wiggins, J. S. (1973). *Personality and prediction: Principles of personality assessment.* Reading, MA: Addison-Wesley.

Zuroff, D. C. (1986). Was Gordon Allport a trait theorist? *Journal of Personality and Social Psychology, 51,* 993–1000.