Psychological Review
1974, Vol. 81, No. 6, pp. 506–520

ON PREDICTING SOME OF THE PEOPLE SOME OF THE TIME:

THE SEARCH FOR CROSS-SITUATIONAL CONSISTENCIES IN BEHAVIOR

DARYL J. BEM[1] AND ANDREA ALLEN

Stanford University

The historically recurring controversy over the existence of cross-situational consistencies in behavior is sustained by the discrepancy between our intuitions, which affirm their existence, and the research literature, which does not. It is argued that the nomothetic assumptions of the traditional research paradigm are incorrect and that by adopting some of the idiographic assumptions employed by our intuitions, higher cross-situational correlation coefficients can be obtained. A study is reported which shows that it is possible to identify on a priori grounds those individuals who will be cross-situationally consistent and those who will not, and it is concluded that not only must personality assessment attend to situations—as has been recently urged—but to persons as well.

Our persistent belief in personality traits, the stubborn assumption that there are pervasive cross-situational consistencies in an individual's behavior, is, quite literally, one of our most ancient convictions:

> Penuriousness is economy carried beyond all measure. A Penurious Man is one who goes to a debtor to ask for his half-obol interest before the end of the month. At a dinner where expenses are shared, he counts the number of cups each person drinks, and he makes a smaller libation to Artemis than anyone. . . . If his wife drops a copper, he moves furniture, beds, chests and hunts in the curtains. . . . [P]enurious men have hair cut short and do not put on their shoes until midday; and when they take their cloak to the fuller they urge him to use plenty of earth so that it will not be spotted so soon [Theophrastus (372–287 B.C.), quoted in Allport, 1937, p. 57].

If this bit of historical personality theorizing has a contemporary ring, it is, in part, because the same underlying assumption of cross-situational consistency is still with us. It is most explicit in trait and type theories of personality, but some variant of it can be discerned in nearly all contemporary formulations. Even psychodynamic theories, which are uniquely competent in dealing with phenotypic inconsistencies in behavior, do so precisely by postulating an underlying genotypic consistency in the personality which rationalizes the apparent contradictions. Our intuitions are even more persuaded. For them the assumption of cross-situational consistency is virtually synonymous with the concept of personality itself. There are few other beliefs about human behavior which are as compellingly self-evident.

But like many other assumptions, the consistency assumption did not fare well during the depression years, when three separate studies with very similar methodologies began to raise serious doubts about its validity. The earliest and best known challenge issued from the extensive multivolume *Studies in the Nature of Character* by Hartshorne and May (1928, 1929; Hartshorne, May, & Shuttleworth, 1930), who found so little consistency among diverse measures of "moral character" in a group of children that they concluded that such traits as deception, helpfulness, cooperativeness, persistence, and self-control are "groups of specific habits rather than general traits." Foreshadowing findings which emerged from hundreds of later studies on scores of personality traits, Hartshorne and May reported that the average intercorrelation of the 23 tests used to construct a "total character score" was a modest +.30.

[1] Requests for reprints should be sent to Daryl J. Bem, Department of Psychology, Stanford University, Stanford, California 94305.

During the same years as the Hartshorne–May inquiry, a less well known but equally troublesome study on extroversion–introversion was published by Theodore Newcomb (1929), who explicitly set out to test the consistency assumption. He kept daily behavioral records on 51 boys at a summer camp for several weeks, recording behaviors in 30 different situations. The behaviors were conceptually organized into 10 separate traits (e.g., volubility versus taciturnity, ascendancy versus submission, etc.) which in turn collectively defined the two personality types of extrovert and introvert.

At the level of specific behaviors, Newcomb found little or no consistency from one situation to another. At the level of trait consistency, the intercorrelations among behaviors composing a given trait averaged only .14, almost identical to the figure obtained from a randomly selected set of behaviors. And finally, there was only a slight tendency for traits to be related to one another as expected by their extrovert–introvert classification.

The third study is in some ways the most damaging of all since it investigated punctuality, a trait one would expect to be much more homogeneous than "moral character" or extroversion–introversion. In this study, Dudycha (1936) made 15,360 observations on over 300 college students, recording each student's time of arrival at 8:00 a.m. classes, commons, appointments, extracurricular activities, vesper services, and entertainments. The mean cross-situational correlation turned out to be +.19, with the highest correlation—between punctuality at entertainments and at commons—reaching .44.

These three studies are virtually unique in that the investigators actually observed behavior in vivo across several situations, a devotion to duty practically unknown in today's literature on the same topic. Meanwhile, the pencil–paper attempts to predict behavior from a trait conception of personality were faring no better, leading Lehmann and Witty (1934) to conclude a review of the literature by saying that "over and over, a battery of tests designed to measure traits such as persistence, or aggressiveness, or honesty, yields results so unreliable and undependable . . . that one is led to question the actual existence of the general traits [p. 490]."

At the same time that the belief in cross-situational consistency was suffering these empirical blows, stimulus–response behaviorism was providing the theoretical argument for the counter belief in the situational specificity of behavior. And with psychologists like Gordon Allport (1937) and Ross Stagner (1937) willing to defend modified trait conceptions of personality against this onslaught, the controversy was a lively one for nearly a decade before receding into the background just prior to World War II (Sanford, 1970).

All of this leads one to appreciate the sense of déjà vu that must currently be affecting psychology's elder statesmen now that the "consistency problem" has suddenly been rediscovered (e.g., Alker, 1972; Allport, 1966; Argyle & Little, 1972; Averill, 1973; D. Bem, 1972; Bowers, 1973; Campus, 1974; Endler, 1973a, 1973b; Endler & Hunt, 1968; Harré & Secord, 1972; Mischel, 1968, 1969, 1973a, 1973b; Moos, 1969; Peterson, 1968; Stagner, 1973; Vale & Vale, 1969; Vernon, 1964; Wachtel, 1973; Wallach & Leggett, 1972).

The major figure in this current round of debate appears to be Walter Mischel (1968), who, after reviewing both past and current research, concludes that the predictive utility of a trait-based approach to personality still remains undemonstrated and that situational specificity of behavior appears to be the rule rather than the exception. Although other contemporary authors have drawn similar conclusions (e.g., Peterson, 1968; Vernon, 1964), it is Mischel who has provoked the most controversy by arguing that the commonly observed +.30 ceiling on cross-situational correlation coefficients probably reflects true behavioral variability rather than imperfect methodology. Since this constitutes a fundamental conceptual challenge, the controversy is once again filling journal pages after a 30-year intermission.

And the stubborn dilemma which sustains this conflict and accounts for its durability still remains unresolved: The sharp dis-

crepancy between our intuitions, which tell us that individuals do in fact display pervasive cross-situational consistencies in their behavior, and the vast empirical literature, which tells us that they do not. Intuitions or research? One of them must be wrong.

ERRORS OF INTUITION AND THE NOMOTHETIC FALLACY OF THE RESEARCH PARADIGM

There are many persuasive reasons for believing that it is our intuitions which are in error (Jones & Nisbett, 1971; Mischel, 1968). First, for example, we hold "implicit personality theories," preconceived notions of what traits and behaviors go with what other traits and behaviors. (See Schneider, 1973, for a recent review.) This leads us not only to generalize beyond our observations and fill in the missing data with "consistent" data of our own manufacture, (e.g., Passini & Norman, 1966) but also to "see" positive correlations which are, in fact, not there (e.g., Chapman & Chapman, 1969; Newcomb, 1929). Moreover, we are biased toward "primacy" effects (e.g., Jones & Goethals, 1971); once we have formed an initial impression of a person, we will perceive inconsistent information about him as more consistent than it deserves to be, assimilating it to our initial judgment.

Second, recent research in attribution theory (e.g., Jones & Davis, 1965; Jones & Harris, 1967; Jones & Nisbett, 1971; Kelley, 1967) has demonstrated that we tend to overestimate the degree to which behavior is caused by traits of the individual and underestimate the degree to which it is caused by external factors. We are therefore willing to generalize about his behavior, extrapolating it to other, unobserved settings in which the situational forces might be quite different.

Third, the set of situations in which we observe most individuals is probably more limited than we realize, both in extent and representativeness. For example, our own presence can frequently evoke a consistent mode of responding in another person (e.g., Kelley & Stahelski, 1970). Accordingly, we are systematically excluded from observing a whole host of situations across which our acquaintances' behaviors are likely to be more variable than they are across the situations in which we do observe them.

Fourth, we probably misconstrue or overgeneralize some of the consistencies which are present. For example, Mischel's (1968) review reveals that the evidence for temporal consistency in behavior is often quite respectable; an individual's behavior is often consistent from one time to another if the situations are similar. Our intuitions may well go from this demonstrable temporal consistency to an unwarranted cross-situational consistency. Moreover, some features of behavior do show cross-situational consistency, such as intellectual ability, cognitive styles, expressive behaviors, and, of course, simple physical appearance. To the extent that these behaviors or cues serve to anchor our inferences about other aspects of behavior—via our implicit personality theories—we will again overgeneralize the degree of cross-situational consistency actually present.

Finally, our language entices us to think of human behavior in trait terms. As Allport and Odbert (1936) reported, there are about 18,000 trait or traitlike terms in our language, nearly five percent of the entire lexicon. In contrast, we have an impoverished and awkward vocabulary for labeling situations.

These, then, are a sample of reasons for thinking that, in the matter of cross-situational consistency, intuitions are wrong, and the research is right. We, however, do not believe it. Despite the compelling impact of these arguments, we still believe that intuitions capture reality more faithfully than does the research. In particular, we believe that there is a basic error in drawing inferences about cross-situational consistency from the traditional research literature in personality, an error which was identified nearly 40 years ago by Gordon Allport (1937). The fallacy resides in the fact that this entire research tradition is predicated upon nomothetic rather than idiographic assumptions about the nature of individual differences. Thus nearly all of the research is based on some variant of the nomothetic assumption that a particular trait dimension

or set of trait dimensions is universally applicable to all persons and that individual differences are to be identified with different locations on those dimensions. For example, the Hartshorne–May study (1928) assumed that an honesty–dishonesty dimension could be used to characterize all of the children in the sample and that the differences among the children could be specified in terms of their *degree* of honesty. A more elaborate version of the same nomothetic assumption can be found in factor-analytic formulations which assume that there is a universal factor structure of personality and that individual differences are to be specified by different points in the factor n-space.

In contrast, Allport's idiographic view emphasized that individuals differ not only in the ways in which traits are related to one another in each person but that they differ also in terms of which traits are even relevant. Thus in commenting upon the fact that Hartshorne and May found lying and cheating to be essentially uncorrelated ($r = .13$), Allport noted that one child may lie because he is afraid of hurting the feelings of the teacher, whereas another may steal pennies in order to buy social acceptance from his peers. For neither of these two children do the behaviors of lying and cheating constitute items on a scale called "honesty," a concept which exists in the head of the investigator, not in the behavior of the children. Accordingly, the low correlations "prove only that children are not consistent *in the same way*, not that they are inconsistent with *themselves* (1937, p. 250)." To put the same objection in slightly different terms, the research will yield the conclusion that a sample of individuals is inconsistent to the degree that their behaviors do not sort into the equivalence class which the investigator imposes by his choice of behaviors and situations to sample.

But there is more. Even if an entire sample of individuals does share the investigator's partitioning of behaviors into the same equivalence class, there is a still more stringent requirement of consistency imposed by the traditional research paradigm: scalability.[2] That is, the sample of individuals must all rank order the "difficulty levels" of the behaviors in the same way.

Consider, for example, the "friendliness" of the second author. She is very friendly to undergraduates in her office, moderately outgoing in a small seminar, and somewhat reserved before a large class. If we can agree that all of these behaviors belong in a common equivalence class labeled "friendliness," she will be judged to be moderately friendly on this trait dimension. She "passes" the "easy" item, has some difficulty with a "harder" item, and "flunks" the "most difficult" item. Note that we do not judge her to be inconsistent any more than we judge a student to be inconsistent when he solves an addition problem but fails a calculus item. We do not do so because their behavior conforms to our a priori ordering of the items in terms of their difficulty levels: their behavior "scales" in the Guttman sense (Scott, 1968; Stouffer, Guttman, Suchman, Lazarsfeld, Star, & Clausen, 1950).

But now consider the "friendliness" of the first author. He, too, passes one item and flunks one item. He is rather formal with undergraduates who appear in his office, moderately outgoing in a small seminar, and open, personable, and friendly before a sea of 300 faces in introductory psychology. But somehow his behavior does not seem describable in terms of the same underlying dimension. He appears not "moderately friendly," but "blatantly inconsistent." And this is because his behaviors do not conform to the a priori Guttman scale which we have implicitly imposed on this equivalence class of behaviors. He passes hard items but flunks easy items.

Now reconsider the traditional research study in which a sample of individuals is assessed on some trait across two or more situations. To the extent that individuals in the sample scale the behaviors differently from one another—as the first and second authors do on "friendliness"—their rankings relative to one another will change from one situation to another. The second author

[2] We are indebted to Stanford colleague Lee Ross for bringing this point to our attention.

will rank first in friendliness in the office encounter; the first author will rank first in the large lecture hall. Under such circumstances, the cross-situational correlation coefficients will plummet toward zero. Only to the extent that all of the individuals in the sample scale the behaviors in the same way will the cross-situational correlations be high.

In summary, then, the traditional trait-based research study will yield evidence of cross-situational consistency only if the individuals in the research sample agree with the investigator's a priori claim that the sampled behaviors and situations belong in a common equivalence class *and* only if the individuals agree among themselves on how to scale those behaviors and situations. The fallacy to which Gordon Allport originally called attention thus becomes clear. The traditional verdict of inconsistency is in no way an inference about individuals; it is a statement about a disagreement between an investigator and a group of individuals and/or a disagreement among the individuals within the group. This fallacy is a direct consequence of the traditional nomothetic assumptions about individual differences. (See a related line of argument by Baldwin, 1946, and McClelland, 1951.)

In contrast to the empirical research, our intuitions operate on idiographic rather than nomothetic assumptions. When we are asked to characterize a friend, we do not invoke some a priori set of fixed dimensions which we apply to everyone. Rather, we permit ourselves to select a small subset of traits which strike us as pertinent and to discard as irrelevant the other 17,993 trait terms in the lexicon. Moreover, we try to compose trait descriptions which conform to the individual's own partitioning of equivalence classes. If John always does his schoolwork early, is meticulous about his personal appearance, and is always punctual, it may well occur to us to describe him as conscientious. But if he is always conscientious about his schoolwork, being negligent in these other areas, we may well describe him as a totally dedicated student, one who has time for little else. The important point is that we are not likely to characterize him as someone who is inconsistently conscientious. That is, we do not first impose a trait term (e.g., conscientious) and then modify it by describing the instances which fail to fall into that equivalence class. Rather, we attempt first to organize his behaviors into rational sets and only then to label them.

We are, moreover, somewhat sensitive to the scaling criterion. We will describe the second author as moderately friendly rather than inconsistent because we recognize the underlying Guttman scale to which her behavior conforms. But when we encounter the first author, whose behavior does not scale according to the recognized "friendliness" dimension, we will typically try to repartition his behaviors before accepting a verdict of inconsistency. Thus the first author is, perhaps, a moderately aloof chap who is a great stage performer. That is, we attempt to construct a new set of equivalence classes which better "capture" the individual's personality. Note also that this intuitive process automatically finessess the problem of situational specificity by embracing in a common equivalence class only those behaviors and situations which cohere for the individual, thereby excluding a priori any maverick behaviors and situations. The trait description is thus fractionated, expanded, contracted, and modified until a best fit of greatest generality and parsimony is achieved. It is only when we fail to discover a set of rationally-scaled equivalence classes which conform to the individual's behavior that a judgment of "inconsistency" is finally rendered. This is the essence of the idiographic approach to personality.

We are not here denying the well-documented biases and illusions which plague our intuitions, nor do we claim that the more formalized idiographic procedures used by clinicians have a better track record in terms of predictive utility than nomothetic ones; they do not (Mischel, 1968). But in terms of the underlying logic and fidelity to reality, we believe that our intuitions are right; the research, wrong.

IDIOGRAPHIC ASSESSMENT AND NOMOTHETIC SCIENCE

The problem with concluding that an idiographic approach represents the path to truth, however, has always been that one is never sure what to do next. To the extent that one accepts psychology's goal as the construction of general nomothetic principles, the idiographic approach appears a scientific dead end, a capitulation to the man-in-the-street view that a science of psychology is impossible because "everybody is different from everybody else." It is this pessimism which seems largely responsible for the fact that the field's respect and admiration for Gordon Allport has never been translated into research programs based upon his conception of personality (cf. Sanford, 1970). Interestingly, the approach of behaviorism has probably discouraged the study of personality differences for much the same reason. To the extent that an individual's behavioral repertoire faithfully reflects the idiosyncratic vagaries of his past reinforcement history, searching for some rational nomothetic basis of personality organization would not appear to be a very hopeful enterprise. As Mischel (1968) notes—not pessimistically, however—the approach of social behavior theory does not

> label the individual with generalized trait terms and stereotypes. . . . Behavioral assessment involves an exploration of the unique or idiographic aspects of the single case, perhaps to a greater extent than any other approach. Social behavior theory recognizes the individuality of each person and of each unique situation [p. 190].

But the impasse is not insurmountable. The use of idiographic assessment procedures did not appear to deter Freud from formulating nomothetic principles of personality organization. Similarly, albeit more modestly, Mischel (1973b) has recently proposed a set of nomothetic principles within the idiographic assumptions of social behavior theory. A third example is provided by George Kelly's (1955) psychology of personal constructs and its associated idiographic assessment procedure, the Role Repertory Test. In fact, it is Kelly's approach which best exemplifies the spirit behind the present arguments. Thus Kelly permits the individual to generate his own traitlike descriptors ("constructs") for characterizing himself and his social world and to determine which behaviors and situations are to be embraced by those descriptors, that is, to determine what Kelly has termed the individual's "range of convenience" for the construct. Note that such an approach could reveal, for example, that an individual who regards himself as extremely conscientious might not consider his casual attitude toward personal hygiene as pertinent to that trait. The fact that the investigator's concept or equivalence class of conscientiousness might include personal hygiene within it is not relevant.

The basic point here is simply that there is no inherent conflict between an idiographic approach to assessment and a nomothetic science of personality, whether one opts for a psychoanalytic orientation, a social learning viewpoint, or a systematization of everyman's trait theory.

It should be clear, however, that idiographic assessment only permits one to predict certain behaviors across certain situations for certain people but not beyond that. Consequently, a conflict does arise if an investigator refuses to relinquish the power to decide which behaviors of which people are to be studied in which situations; the logic of idiographic assessment requires that the individual himself must be given this power, whereas the particular concerns of the investigator may require these decisions to be fixed parameters.

Consider, for example, the researcher who wishes to study, say, need for achievement in a particular setting in a particular population. No matter how persuasive he finds our arguments for the merits of idiographic assessment, he is simply not interested in studying a different set of personality variables in each individual. But, on the other hand, our arguments imply that need for achievement may not even *be* a trait dimension which usefully characterizes many of the individuals in the sample. As his low validity coefficients will attest, those individuals will contribute only noise to his investigation. The dilemma is real. If our arguments here are sound, one simply can-

not, in principle, ever do any better than predicting some of the people some of the time. It is an idiographic fact of life.

Our advice to such an investigator, then, follows directly: Find those people. Separate those individuals who are cross-situationally consistent on the trait dimension and throw the others out, for by definition, only the behavior of consistent individuals can be meaningfully characterized by the investigator's construct; only their behaviors can be partitioned into the equivalence class under investigation. Perhaps a statistical metaphor will make this proposal seem less illegitimate: Unless an individual's variance on a particular trait dimension is small, it makes no sense to attach psychological significance to his mean on that dimension.

We submit that even this token gesture toward more idiographic assessment has its rewards. First of all, one may obtain valuable knowledge about the trait dimension itself; it could be useful (as well as humbling) to discover why, which, and how many individuals fail to share the investigator's partitioning of the world into his favorite equivalence class. But perhaps even better, we believe that the rewards for this small idiographic commitment can even be paid in the sacred coin of the realm: bigger correlation coefficients! The following demonstration illustrates the point.

A PRIORI ASSESSMENT OF CROSS-SITUATIONAL CONSISTENCY [3]

Our purpose in this study was to test whether or not individuals can be divided on the basis of self-report into those who are cross-situationally consistent on a particular trait and those who are not. Our hypothesis is straightforward: Individuals who identify themselves as consistent on a particular trait dimension will in fact be more consistent cross-situationally than those who identify themselves as highly variable. In population terms, the cross-situational correlation coefficients of the self-identified low-variability group should be significantly higher than the coefficients of the high-variability group. In order to add a bit more persuasive elegance to the study, we tested this hypothesis twice on the same population of subjects using two orthogonal personality traits, friendliness and conscientiousness.

Method

As part of a questionnaire entitled the Cross-Situation Behavior Survey (CSBS), all students in Stanford's introductory psychology course were asked to assess themselves on several trait dimensions, including friendliness and conscientiousness. On each dimension, the individual was asked to rate both his overall level and his variability. For example, on the friendliness dimension, he was asked, "In general, how friendly and outgoing are you?" and "How much do you vary from one situation to another in how friendly and outgoing you are?" Parallel pairs of questions were asked about conscientiousness and other traits. Responses were obtained on a seven-point scale which ranged from "not at all' to "extremely." It will be noted that these questions thus permit the individual to employ his own concept of the trait dimension, to average across the situations he sees as pertinent and to ignore situations he sees as irrelevant. Accordingly, these global self-ratings will be successful in predicting behavior only to the extent that the individual's definition of a trait dimension coincides with the definition we will necessarily be imposing by our selection of situations to sample.

Using the same seven-point response scale, we also obtained each individual's self-ratings on specific behavior–situation items for each trait. For example, the CSBS included a 24-item scale which assessed the trait of friendliness in specific situations (e.g., "When in a store, how likely are you to strike up a conversation with a sales clerk?") and a 23-item conscientiousness scale (e.g., "How carefully do you double-check your term papers for typing or spelling errors?"). Thus if the global self-ratings can be seen as reflecting the individual's own definitions of the trait dimensions, then these CSBS scales can be viewed as reflecting the investigators' conception of these dimensions. The internal reliabilities (coefficient alpha) of the two trait scales were .91 and .84 for friend-

[3] We are grateful to many people for help in executing this study, particularly to Fred Bart Astor who served as overall supervisor and research assistant and to Margaret Bond who transformed raw observations into meaningful data. For their multiple roles as trained observers, experimental confederates, and general research assistants, we are grateful to John Backes, Margaret Bond, Kathleen Chiappori, Tom Deremigio, Gowen Roper, Jeremy Rosenblum, and Ann Scholey. Above all, however, we are grateful to our subjects, who were genuine partners in this research effort. In return for their cooperation, we attempted to keep them as fully informed of our procedures and rationale as the design would permit.

liness and conscientiousness, respectively, and the correlation between them was +.13.

Cross-situational assessment. From the introductory psychology course, 32 male and 32 female students were recruited as subjects. In addition to the initial testing session in which all students participated, the subjects were seen on three separate occasions, and they also signed release forms giving us permission to obtain ratings on them from their parents and one of their close peers, usually a roommate. From these various sessions, the following six friendliness variables and seven conscientiousness variables were derived.

Friendliness: (1) *Self-report;* (2) *Mother's Report* (3) *Father's Report* (4) *Peer's Report:* Each of these four judges provided us with an independent assessment of the individual's friendliness by rating him on the global friendliness item and the 24-item CSBS friendliness scale. For each judge, these two measures were combined into a single score. (5) *Group Discussion:* Each individual was observed as he participated in a group discussion with three other subjects of the same sex. A measure of each individual's friendliness in the group was derived from the frequency and duration of his vocalizations and the group's postdiscussion rating of his friendliness. (6) *Spontaneous Friendliness:* Each individual was observed as he waited in a waiting room with an experimental confederate, and a measure of spontaneous friendliness was derived from his "latency" in initiating conversation.

Conscientiousness: (1) *Self-Report;* (2) *Mother's Report;* (3) *Father's Report;* (4) *Peer's Report:* As with friendliness, each of these four judges provided us with an independent assessment of the individual's conscientiousness by rating him on the global conscientiousness item and the 23-item CSBS conscientiousness scale. (5) *Returning Evaluations:* During the quarter, each individual received four evaluation forms by mail as part of an ongoing assessment of the introductory psychology course. Each form asked him to evaluate one of the course lectures and to return the form anonymously prior to the subsequent class period. (The forms for our subjects were numerically coded.) Measures of his promptness in returning each of the four forms were combined into a single index. (6) *Course Readings:* Each of the course evaluation forms described above asked the individual to check off the course readings he had been able to complete up to that time, providing us with four separate reports of his conscientiousness in school work. (7) *Neatness:* The neatness and cleanliness of each individual's hair and clothing were rated on two separate occasions by three independent judges and his living quarters were rated on nine aspects of neatness during a surprise visit paid during the last week of the quarter. These several observations were combined into an overall neatness score.

Finally, it should be noted that experimental assistants and observers were all blind with respect to the individual's scores on the trait dimensions, and no observer made more than one observation for each trait on the same individual.

Results

The first step in the analysis of results was to classify each individual on a priori grounds as a low-variability or a high-variability subject in a way that would not be confounded with his actual position on the trait dimension. Accordingly, for each trait, a subject was first classified into one of seven subgroups on the basis of his response to the question "In general, how friendly and outgoing [conscientious] are you?". Then, on the basis of his response to the question "How much do you vary from one situation to another in how friendly and outgoing [conscientious] you are?", he was designated as a low-variability or a high-variability subject, respectively, depending upon whether he was below or above the median among the same-sex subjects at the same point on the trait scale. Thus low and high variability were redefined at each of the seven points on the global trait scale in order to partial out any relationship between an individual's self-rated variability and his self-rated position on the trait dimension.

In order to assess each individual's cross-situational consistency for each trait, we converted each of the 13 variables to a standard T score with a mean of 50 and a standard deviation of 10 across the 64 subjects. We then calculated each individual's standard deviation across the six friendliness variables and across the seven conscientiousness variables. These two standard deviations thus reflect the individual's cross-situational variability on friendliness and conscientiousness, respectively; the larger the standard deviation, the more variable he is across situations.

Friendliness. With respect to the friendliness dimension, our hypothesis was confirmed. Individuals who indicate that they do not vary much from one situation to another do, in fact, display significantly less variability across situations than do those who say they do vary (6.42 versus 7.90; $t = 2.34$, $p < .02$, one-tailed test). Moreover, an individual's self-rated friendliness per se

TABLE 1

INTERCORRELATIONS AMONG THE SIX FRIENDLINESS VARIABLES FOR LOW- AND HIGH-VARIABILITY SUBJECTS

Low / High	Self-report	Mother's report	Father's report	Peer's report	Group discussion	Spontaneous friendliness	All variables
1. Self-Report							
2. Mother's Report	.61 / .52						
3. Father's Report	.48 / .24	.75 / .28					
4. Peer's Report	.62 / .56	.71 / .40	.50 / .34				
5. Group Discussion	.52 / .59	.34 / .41	.50 / .13	.45 / .39			
6. Spontaneous Friendliness	.61 / −.06	.46 / −.18	.69 / −.20	.39 / .09	.73 / .30		
Mean correlations	.57 / .39	.59 / .30	.60 / .16	.54 / .37	.52 / .37	.59 / .01	.57 / .27

is not related to his cross-situational variability; in particular, individuals in the lowest, middle, and highest thirds of the distribution on the self-rated friendliness scale did not differ from one another in their cross-situational variability, $F\ (2,\ 61) = 1.10$, *ns*.

Table 1 shows how the differential cross-situational consistency of low- and high-variability individuals translates into cross-situational predictability in correlational terms. The intercorrelations of the six variables for the 32 low-variability subjects are shown above the diagonal; intercorrelations for the 32 high variability subjects, below the diagonal. The bottom row of Table 1 serves to summarize the matrix by showing the mean correlation between the column variable and the remaining five variables for the two groups separately.[4]

As Table 1 shows, 13 of the 15 intercorrelations are higher for low-variability subjects than they are for high-variability subjects, six of them significantly so ($p < .05$, one-tailed test). The mean intercorrelation among all the variables is +.57 for the low-variability group and +.27 for the high-variability group. Note also that the predicted effect is quite general across different pairs of situations. For example, Mother's Report and Father's Report, two measurements that one would expect to be highly similar and "contaminated" by one another, show a correlation of +.75 for the low-variability group but only +.28 for the high-variability group ($p < .005$, one tailed test); similarly, Group Discussion and Spontaneous Friendliness, the two methodologically independent behavioral observations, show a correlation of +.73 for the low-variability group but only +.30 for the high-variability group ($p < .009$, one-tailed test). Thus, not only have our expectations been confirmed, but the magic +.30 barrier appears to have been penetrated.

It will be noted that the "moderating variable" in this analysis, the variable which separates the population into groups which are differentially predictable, is the individual's response to the single question "How much do you vary from one situation to another in how friendly and outgoing you are?" To see whether such an item could enhance the utility of a standard personality inventory as well, we computed the cor-

[4] All calculations and analyses involving correlation coefficients are actually performed on their *Z*-transforms.

relations between the six friendliness variables and the extraversion–introversion scale of the Eysenck Personality Inventory (Eysenck & Eysenck, 1968), which all of our subjects had taken at the initial testing session. Using the same criterion for low and high variability as before, Table 2 displays the correlations for the two groups separately. It is seen that the extraversion–introversion scale of the Eysenck Personality Inventory does indeed have greater predictive utility for self-identified low-variability subjects than for high-variability subjects ($t = 3.96$, $p < .01$). Moreover, the effect remains significant even if the methodologically similar Self-Report variable is removed from the analysis ($t = 3.22$, $p < .025$).

TABLE 2

CORRELATIONS BETWEEN EYSENCK'S EXTRAVERSION–INTROVERSION SCALE AND SIX FRIENDLINESS VARIABLES AS A FUNCTION OF SELF-RATED VARIABILITY

Extraversion versus:	Self-rated variability	
	Low	High
Self-Report	.77	.65
Mother's Report	.54	.37
Father's Report	.26	.24
Peer's Report	.71	.41
Group Discussion	.34	.18
Spontaneous Friendliness	.25	−.12
Mean correlation[a]	.51	.31
Mean correlation (omitting self-report)[b]	.44	.22

[a] Correlated $t = 3.96$, $p < .01$, one-tailed test.
[b] Correlated $t = 3.22$, $p < .025$, one-tailed test.

Conscientiousness. As we noted earlier, the global self-rating items used to classify our subjects permit the individual to utilize his own definition of the trait, to implicitly average across behaviors he regards as pertinent and to ignore all others. As we pointed out, this seemed likely to yield our predicted results only to the extent that the subjects' trait definitions coincided with our own. For the trait of friendliness, an adequate commonality of definition was achieved. For example, the correlation between the individual's global self-rating of his friendliness and his mean score on our CSBS friendliness items was +.84. The corresponding correlation for the conscientiousness trait, however, was significantly lower, $r = .62$; $Z_{\text{difference}} = 2.74$, $p < .006$, two-tailed test, implying that the trait term "conscientiousness" is more likely to denote different equivalence classes of behaviors for different individuals than is the trait term "friendliness." Not unexpectedly, then, we were not able to replicate the friendliness findings for the conscientiousness trait when we employed our subjects' global self-ratings as the classification variables.

Accordingly, we turned to *our* definition of conscientiousness as the basis for subject classification, designating each individual as low or high variability as a function of his variance on the CSBS conscientiousness scale. In particular, we calculated each individual's variance across the 23 conscientiousness items and divided it by his variance across all 86 items of the questionnaire. This "ipsatized" variance index not only corrects for the individual's tendency to respond consistently or inconsistently to CSBS items irrespective of their content, but it also has a more conceptual interpretation as well. It reflects the degree to which an individual "extracts" the particular trait-scale items from the total pool of items and "clusters" them into an equivalence class. Statistically, the ipsatized variance is like an inverted F ratio, representing the ratio of two variances which assumes a value of zero if the individual responds identically to each item on the trait scale and a value of one if he does not "cluster" the items on a trait scale at all.[5]

Subjects were first formed into matched pairs on the basis of their CSBS conscientiousness scores, and then each individual was classified as low or high variability, respectively, depending upon whether his ipsatized variance was lower or higher than that of his matched partner. It will be recognized that this again serves to partial out any relationship between the individual's variability and his location on the trait dimension. As before, the individual's

[5] In 1961, Berdie used an intraindividual variance measure as a moderating variable for predicting mathematical aptitude. (See also Campus, 1974, and Fiske and Rice, 1955.)

TABLE 3

INTERCORRELATIONS AMONG THE SEVEN CONSCIENTIOUSNESS VARIABLES FOR LOW- AND HIGH-VARIABILITY SUBJECTS

Low / High	Self-report	Mother's report	Father's report	Peer's report	Returning evaluations	Course readings	Neatness	All variables
1. Self Report								
2. Mother's Report	.47 / .27							
3. Father's Report	.31 / .23	.83 / .46						
4. Peer's Report	.44 / .25	.50 / .04	.38 / .49					
5. Returning Evaluations	.43 / −.12	.60 / .03	.42 / .10	.49 / .11				
6. Course Readings	.40 / −.07	.20 / −.31	.31 / −.24	.61 / −.16	−.01 / .18			
7. Neatness	.31 / .55	.13 / .39	.11 / .43	.23 / .45	−.01 / −.14	−.11 / −.61		
Mean correlations	.40 / .20	.50 / .15	.43 / .26	.45 / .21	.33 / .03	.25 / −.22	.12 / .19	.36 / .12
Mean correlations (omitting Neatness)	.41 / .11	.56 / .10	.49 / .22	.49 / .16	.40 / .06	.32 / −.12	—	.45 / .09

cross-situational variability is assessed by his standard deviation across the several situations.

With this method of classification, the results confirm our hypothesis, paralleling exactly our findings for the friendliness dimension. Thus low-variability subjects were significantly less variable across situations than high-variability subjects (7.46 versus 8.89, correlated $t = 2.80$, $p < .005$, one-tailed test). And again, an individual's standing on the trait itself was not related to his cross-situational variability; individuals in different thirds of the distribution on the CSBS conscientiousness scale did not differ from one another in their cross-situational variability, $F(2,61) < 1$, *ns*.[6]

Table 3 translates the differential cross-situational consistency of low- and high-vari-

[6] We believe that, like conscientiousness, most traits would not attain the degree of definitional consensus between subject and investigator displayed by the friendliness trait. Accordingly, we believe that the ipsatized variance index, rather than the single item self-rating of variability, will prove to be the more promising candidate for the moderating variable in any future work. Moreover, the ipsatized variance index can be calculated for any set of questionnaire items the investigator chooses even when the layman has no trait term for labeling the potential equivalence class so defined.

ability individuals into correlational terms. As before, correlations for the 32-low variability subjects are found above the diagonal, the 32 high-variability subjects, below the diagonal.

It is seen in Table 3 that 15 of the 21 intercorrelations are higher for low-variability subjects than they are for high-variability subjects, with 9 of them significantly so ($p < .05$, one-tailed test). As the summary rows of the table reveal, only the Neatness variable fails to conform to our hypothesis accounting for all but two of the reversals in the matrix. The bottom row of the table shows the mean intercorrelations obtained when this variable is omitted.

But it is precisely the Neatness variable which illustrates the main point of this article. As an inspection of the correlation matrix reveals, it is only we, the investigators, who think that school-related conscientiousness (Returning Evaluations and Course Readings) and personal neatness ought to belong in the same equivalence class. Our subjects do not. Our low-variability subjects find them to be orthogonal ($-.01$ and $-.11$), and our high-variability subjects apparently have time to do their schoolwork *or* to keep things neat and clean, but not both ($r = -.61$).

The judgments by self, parents, and peers are also interesting in this regard. An inspection of the correlations in the matrix reveals that judges of low-variability subjects appear to be responding to both kinds of conscientiousness almost equally, with some bias toward greater attention to school-related conscientiousness. But judges of high-variability subjects seem to be ignoring school-related conscientiousness and responding primarily to the inversely-related personal neatness. It is this latter pattern which causes the Neatness variable to violate our hypothesis that low-variability subjects would have the higher correlations.

The moral of all this, of course, is that we need to move even further toward idiographic assessment. In this demonstration, we have relinquished the presumption that all traits are relevant to all people but stubbornly retained the right to dictate which behaviors and situations shall constitute the trait itself. When an investigator is willing to release this degree of freedom as well, his validity coefficients will reward him with an appropriate increment in magnitude. In the present study, for example, we might have asked the subjects to rate the several CSBS items for their relevance to the various trait dimensions. In this way, we might have discovered a priori those subjects who do not share our personal delusion that the conscientious person does his schoolwork *and* attends to his personal neatness.

PREDICTIVE UTILITY OF TRAITS AND SITUATIONS

We have argued in this article that it is not possible, in principle, to do any better than predicting some of the people some of the time. Furthermore, our arguments would seem to imply that an investigator must simply abandon the highly variable individual since the trait under investigation has no predictive utility for him. But this is not always true. As Mischel (1968, 1973b) has persuasively argued, variability is not synonymous with either capriciousness or unpredictability. Indeed, an individual's cross-situational variability may well be the mark of a highly refined "discriminative facility" (Mischel, 1973b), the ability to respond appropriately to subtle changes in situational contingencies. Although such an individual cannot be predicted from a knowledge of his standing on a personality trait, he may be precisely the individual who is most predictable from a knowledge of the situation. In short, if some of the people can be predicted some of the time from personality traits, then some of the people can be predicted some of the time from situational variables.

This point is nicely illustrated in recent work on sex roles by S. Bem (1974, in press). Whereas previous research in this area has been concerned either with the sex-typed masculine males and feminine females or, occasionally, with the sex-reversed feminine males and masculine females, S. Bem has constructed a sex-role inventory which permits her to identify "androgynous" individuals as well, individuals who attribute both masculine and feminine characteristics

to themselves in about equal amounts without regard for their sex-role connotations. Thus, with respect to either of the two trait terms masculinity or femininity, androgynous individuals are the "high variability" subjects; neither trait has predictive utility for them. And as hypothesized, Bem finds that androgynous individuals of both sexes vary their behavior cross-situationally so that they are able to "do well" at both masculine and feminine tasks and behaviors, whereas sex-typed individuals do not do well in situations or settings which call for behaviors incongruent with their self-described sex roles. For example, only the androgynous subjects showed both "masculine" independence in an Asch conformity experiment as well as "feminine" nurturance when given the opportunity to play with a baby kitten (S. Bem, in press). There are, in short, some occasions when one can predict some of the high-variability people some of the time.

It should be clear from this discussion that the position we have argued in this article cannot be characterized as opposing either side of the debate between those who believe that behavior is consistent across situations and those who believe that behavior is situationally specific. The shift to idiographic assumptions about the nature of individual differences dissolves this false dichotomy and permits one to believe in both propositions. As noted early in this article, the actual cleavage is between nomothetic and idiographic criteria for consistency and inconsistency.

Similarly, it should be clear that we have not been arguing an "anti-Mischel" position. Both Mischel and we agree that the nomothetic assumptions of the traditional approach in personality virtually guarantee a +.30 ceiling on validity coefficients and, hence, that trait-based approaches predicated on such assumptions will continue to fail the test of predictive utility. Mischel and we agree that only an idiographic approach can break through this predictive barrier. Mischel and we agree that the classification of situations must be an integral part of any assessment procedure; moreover, we agree that such classification will have to be in terms of the individual's own phenomenology, not the investigator's (Mischel, 1973b), a suggestion that is bound to increase further the déjà vu of any psychologist old enough to remember Kurt Lewin (1935). It is true that Mischel and we have chosen different conceptual languages in which to express these points, and future divergences will surely emerge as a consequence, but the two formulations are far more similar in their basic assumptions than their formal appearances would suggest.

The failure of traditional assessment procedures and the belief that person–situation interactions will account for most of the psychologically interesting variance in behavior have led several recent writers to emphasize that personality assessment must begin to attend seriously to situations. We agree. We have merely chosen to emphasize the perfectly symmetric, but perhaps more subtle, point that personality assessment must also begin to attend seriously to persons.

REFERENCES

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

Allport, G. W. *Personality: A psychological interpretation.* New York: Holt, 1937.

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

Allport, G. W., & Odbert, H. S. Trait-names: A psycholexical study. *Psychological Monographs*, 1936, **47** (1, Whole No. 211).

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

Averill, J. R. The dis-position of psychological dispositions. *Journal of Experimental Research in Personality*, 1973, **6**, 275–282.

Baldwin, A. L. The study of individual personality by means of the intraindividual correlation. *Journal of Personality*, 1946, **14**, 151–168.

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

Bem, S. L. The measurement of psychological androgyny. *Journal of Consulting and Clinical Psychology*, 1974, **42**, 155–162.

Bem, S. L. Sex-role adaptability: One consequence of psychological androgyny. *Journal of Personality and Social Psychology*, in press.

Berdie, R. F. Intra-individual variability and predictability. *Educational and Psychological Measurement*, 1961, **21**, 663–676.

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

Campus, N. Transituational consistency as a dimension of personality. *Journal of Personality and Social Psychology*, 1974, **29**, 593–600.

Chapman, L. J., & Chapman, J. P. Illusory correlations as an obstacle to the use of valid psycho-diagnostic signs. *Journal of Abnormal Psychology*, 1969, **74**, 271–280.

Dudycha, G. J. An objective study of punctuality in relation to personality and achievement. *Archives of Psychology*, 1936, **204**, 1–319.

Endler, N. S. The case for person-situation interactions. Paper presented at the meeting of the American Psychological Association, Montreal, August 1973. (a)

Endler, N. S. The person versus the situation—A pseudo issue? A response to Alker. *Journal of Personality*, 1973, **41**, 287–303. (b)

Endler, N. S., & Hunt, J. McV. S-R inventories of hostility and comparisons of the proportions of variance from persons, responses, and situations for hostility and anxiousness. *Journal of Personality and Social Psychology*, 1968, **9**, 309–315.

Eysenck, H. J., & Eysenck, S. B. G. *Manual for the Eysenck Personality Inventory*. San Diego: Educational and Industrial Testing Service, 1968.

Fiske, D. W., & Rice, L. Intra-individual response variability. *Psychological Bulletin*, 1955, **52**, 217–250.

Harré, R., & Secord, P. F. *The explanation of social behavior*. Oxford: Basil, Blackwell, & Mott, 1972.

Hartshorne, H., & May, M. A. *Studies in the nature of character*. Vol. 1. *Studies in deceit*. New York: Macmillan, 1928.

Hartshorne, H., & May, M. A. *Studies in the nature of character*. Vol. 2. *Studies in service and self-control*. New York: Macmillan, 1929.

Hartshorne, H., May, M. A., & Shuttleworth, F. K. *Studies in the nature of character*. Vol. 3. *Studies in the organization of character*. New York: Macmillan, 1930.

Jones, E. E., & Davis, K. E. From acts to dispositions: The attribution process in person perception. In L. Berkowitz (Ed.), *Advances in experimental social psychology*. Vol. 2. New York: Academic Press, 1965.

Jones, E. E., & Goethals, G. R. *Order effects in impression formation: Attribution context and the nature of the entity*. New York: General Learning Press, 1971.

Jones, E. E., & Harris, V. A. The attribution of attitudes. *Journal of Experimental Social Psychology*, 1967, **3**, 1–24.

Jones, E. E., & Nisbett, R. E. *The actor and observer: Divergent perceptions of the causes of behavior*. New York: General Learning Press, 1971.

Kelley, H. H. Attribution theory in social psychology. In D. Levine (Ed.), *Nebraska Symposium on Motivation: 1967*. Lincoln: University of Nebraska Press, 1967.

Kelley, H. H., & Stahelski, A. J. Social interaction basis of cooperators' and competitors' beliefs about others. *Journal of Personality and Social Psychology*, 1970, **16**, 66–91.

Kelly, G. A. *The psychology of personal constructs*. New York: Norton, 1955, 2 vols.

Lehmann, H. C., & Witty, P. A. Faculty psychology and personality traits. *American Journal of Psychology*, 1934, **44**, 490.

Lewin, K. *A dynamic theory of personality*. New York: McGraw-Hill, 1935.

McClelland, D. C. *Personality*. New York: William Sloane Associates, 1951.

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

Mischel, W., Continuity and change in personality. *American Psychologist*, 1969, **24**, 1012–1018.

Mischel, W. On the empirical dilemmas of psychodynamic approaches: Issues and alternatives. *Journal of Abnormal Psychology*, 1973, **82**, 335–344. (a)

Mischel, W. Toward a cognitive social learning reconceptualization of personality. *Psychological Review*, 1973, **80**, 252–283. (b)

Moos, R. H. Sources of variance in responses to questionnaires and in behavior. *Journal of Abnormal Psychology*, 1969, **74**, 405–412.

Newcomb, T. M. *Consistency of certain extrovert–introvert behavior patterns in 51 problem boys*. New York: Columbia University, Teachers College, Bureau of Publications, 1929.

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

Peterson, D. R. *The clinical study of social behavior*. New York: Appleton-Century-Crofts, 1968.

Sanford, N. *Issues in personality theory*. San Francisco: Jossey-Bass, 1970.

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

Scott, W. A. Attitude measurement. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology*. Vol. 2. Menlo Park, Calif.: Addison-Wesley, 1968.

Stagner, R. *Psychology of personality*. New York: McGraw-Hill, 1937.

Stagner, R. Traits are relevant—logical and empirical analysis. Paper presented at the meeting of the American Psychological Association, Montreal, August 1973.

Stouffer, S. A., Guttman, L. Suchman, E. A., Lazarsfeld, P. F., Star, S. A., & Clausen, J. A. *Studies in social psychology in World War II*.

Vol. 4. *Measurement and prediction.* Princeton: Princeton University Press, 1950.
Vale, J. R., & Vale, G. R. Individual difference and general laws in psychology: A reconciliation. *American Psychologist,* 1969, **24,** 1093–1108.
Vernon, P. E. *Personality assessment: A critical survey.* New York: Wiley, 1964.
Wachtel, P. Psychodynamics, behavior therapy, and the implacable experimenter: An inquiry into the consistency of personality. *Journal of Abnormal Psychology,* 1973, **82,** 324–334.
Wallach, M. A., & Leggett, M. I. Testing the hypothesis that a person will be consistent: Stylistic consistency versus situational specificity in size of children's drawing. *Journal of Personality,* 1972, **40,** 309–330.

(Received January 21, 1974)

NOTICE TO APA AUTHORS

The American Psychological Association announces publication of the second edition of the *Publication Manual* in August 1974. This new edition, which supersedes the 1967 Manual, will be adopted by the 14 APA journals in 1975.

The new Manual is more comprehensive than the previous edition. It updates APA publication policies and procedures and incorporates changes in editorial practice since 1967. For instance, APA now sends many authors their copy-edited manuscripts for review before they are set into type, and some editors now use blind review procedures. The second edition also includes up-to-date statements of the coverage of each APA journal including the *Journal of Experimental Psychology* which will be published in four separate sections in 1975.

The new *Publication Manual* initiates several changes in APA style. These changes are announced in the August 1974 *American Psychologist* and will be introduced in the APA journals in January 1975. During the period of transition to the new style, authors should note that (*a*) all manuscripts published in 1974 will be copy-edited according to the 1967 Manual, (*b*) manuscripts accepted in 1974 and published in 1975 will be copy-edited to conform to the new Manual. Starting in 1975, accepted manuscripts that depart significantly from the Manual will be returned to authors for correction.

Authors will be encouraged by the changes in the second edition. The new APA style simplifies reference forms; eliminates unnecessary underlines, brackets, and other devices; supports appropriate use of "I" and "we"; and generally clarifies typing requirements. Material is arranged for maximum convenience to authors and typists, and all sections are cross-referenced and indexed.

The new *Publication Manual* is available after August 1 for $3. Send orders to APA Publication Sales, 1200 Seventeenth Street, N.W., Washington, D. C. 20036. Orders of $15 or less must include payment unless they are submitted on institutional purchase order forms.