Journal of Personality and Social Psychology
1991, Vol. 60, No. 5, 773-794

Copyright 1991 by the American Psychological Association, Inc.
0022-3514/91/$3.00

Explorations in Behavioral Consistency: Properties of Persons, Situations, and Behaviors

David C. Funder and C. Randall Colvin
University of California, Riverside

In a study exploring the cross-situational consistency of behavior, 140 undergraduate Ss were videotaped in each of 3 laboratory settings, and personality descriptions of these Ss were obtained from friends and acquaintances. Analyses focused on the degree to which Ss maintained consistent patterns of behavior across laboratory settings and between these settings and daily life. The following conclusions were reached: (a) Behavior can exhibit impressive consistency at the level of psychological meaning, (b) psychological properties of situations can be detected from their behavioral effects, (c) cross-situational consistency and discriminativeness are independent, and (d) some behaviors are more consistent than others.

A young man and woman who have never met before arrive for a psychology experiment. The experimenter immediately ushers them into a small room that contains a couch, a TV camera, and a video recorder. "Have a seat," says the experimenter, aiming the camera at the couch, "I'll be back in a few minutes." The experimenter activates the video recorder and departs.

What happens next varies widely. Some people, seemingly oblivious to the camera and to the strangeness of this situation, plunge immediately into friendly and relaxed conversation about sports, current events, vacation plans, college majors, and grades. The experimenter's return is an unwelcome intrusion. Others look awkwardly at the floor, shuffle their feet, perhaps exchange names, and lapse into unhappy silence. The experimenter's return is greeted with relief. Still others put on a bit of a show for the camera, making funny faces and addressing various remarks directly to the (unseen) researchers. Some people dominate the interaction, some appear fearful, and others seem merely apathetic.

After a few weeks, these people are recruited for a second experiment and again are paired with partners they have never met before. The videotape rolls, and a wide range of individual differences in behavior is again seen. A few minutes later, the experimenter asks them to engage in a debate about capital punishment. They vary in their responses to this situation as well. Some debate with vigor, and others could not seem to care less; some are confident, and some are nervous, and so forth.

The explorations in the present article begin with two deceptively simple questions:

1. What does the behavior of a subject in one of these three situations have to do with his or her behavior in the others? For instance, is a subject who is relatively fearful at one time also relatively fearful at another? Does the same subject who dominates one of these interactions also tend to dominate the other two?

2. What does the behavior of subjects in these situations have to do with their behavior in real life? For instance, are people who show the most intellectual interest during the debate on capital punishment the same ones who generally exhibit the most intellectual interest back at the dorm? Are subjects who express the most emotion in these three situations also the most emotionally expressive in daily interactions with their friends and acquaintances?

Research on Behavioral Consistency

The principal job of personality psychology is to account for the organization of action, thought, and experience in the life of each person. Historically, a prominent approach to this task has been to examine the patterns of behavior that people manifest over time and across situations (e.g., Allport, 1937). This approach provides a basis for personality trait constructs and the technology and practice of personality assessment (Wiggins, 1973).

The assumption of behavioral consistency has been a useful heuristic for research. A vast and venerable literature has examined subjects across a large variety of situations and sometimes over many years. Examples include classic "assessment center" studies in the tradition of Henry Murray (1938) and the Institute of Personality Assessment and Research in Berkeley (e.g., McKinnon, 1962). Other examples are classic studies of personality over time such as the Kelly Longitudinal Study (Conley, 1985; E. L. Kelly, 1955), the Terman Study (e.g., Terman & Oden, 1947), the Oakland Growth Study (Block, 1971), the Block project (Block & Block, 1980), and longitudinal studies of managers such as the one conducted at AT&T (Howard & Bray, 1988). The result has been a wealth of knowledge concerning, among other things, the personality dynamics of creative architects, long-range effects of various styles of parenting and

The research reported in this article was supported by National Institute of Mental Health Grants MH40808 and MH42427 to David C. Funder. We are grateful for comments offered by Robert Hogan, John Kihlstrom, Jerry Wiggins, and three anonymous reviewers on earlier drafts but remain solely responsible for the views expressed herein.

Correspondence concerning this article should be addressed to David C. Funder, Department of Psychology, University of California, Riverside, California 92521.

teaching, and the interaction between personality and success within different organizational environments.

This research has been more often concerned, however, with the *antecedents* or *consequences* of consistent behavior patterns than with the nature of behavioral consistency itself. For the most part, empirical studies that have addressed behavioral consistency as a topic in its own right have reached rather pessimistic conclusions—pessimistic, at least, from the point of view of anyone who regards behavioral consistency as an essential attribute of personality. Hartshorne and May (1928) observed the cheating behavior of children at various games, found that a child who cheated at one did not necessarily cheat at another, and concluded not only that the trait of honesty did not exist, but that moral behavior, rather than stemming from the consistent character of people, was simply a product of specific situations (but see Peck & Havighurst, 1960). Sixty years later this study continues to garner frequent citations.

Somewhat more recently, Mischel and Peake (1982) reported a complex study of behavioral consistency that yielded a large number of findings, a typical one being that an aggregated measure of the thoroughness of students' class notes correlated only −.03 with an aggregated measure of their punctuality to lectures (p. 735)—both behaviors being putative manifestations of conscientiousness. This single study evoked so much interest that it managed to spawn a small literature of its own (Bem, 1983; Epstein, 1983; Funder, 1983; Jackson & Paunonen, 1985; Mischel & Peake, 1983). Its principal conclusion will sound highly familiar to anyone who has followed the research literature on personality in recent years: "It is . . . clear from these results that behavior is . . . highly discriminative and . . . broad cross-situational consistencies remain elusive" (p. 735).

A recent review of the personality literature concluded that behavior probably exhibits more consistency than the quotation in the previous paragraph might seem to imply, according to existing evidence concerning interjudge agreement in personality ratings and behavioral correlates of personality (Kenrick & Funder, 1988). Nonetheless, after more than two decades of debate, much is still unknown about the circumstances under which behavior will be consistent from one laboratory situation to another and between laboratory situations and real life. A reexamination of these questions raises four fundamental issues, which are discussed in the next four subsections.

Behavioral Consistency and Level of Analysis

The usual conceptualization of behavioral consistency has been as something that is manifest at the level of concrete acts (e.g., Bem & Allen, 1974; Mischel & Peake, 1983). This viewpoint treats the measurement of consistency as a simple matter of counting: If the number of times a person does something in one situation is perfectly predictable from the number of times he or she does something equivalent in another situation, then his or her behavior is consistent. If this number is not so predictable, then his or her behavior is inconsistent. For instance, as already mentioned, in their investigation of honesty Hartshorne and May (1928) ascertained whether a child cheated at each of several games. To examine the consistency of conscientiousness, Mischel and Peake (1982) counted specific behaviors such as class attendance, assignment punctuality, and completion of class readings.

The analysis of concrete acts was a reasonable place for examination of behavioral consistency to begin. However, two decades of subsequent research have revealed that consistency of this sort is quite rare if it exists at all. This may be because the initial conceptualization was too simple. A child might refrain from cheating in one setting out of fear of being caught but in another because cheating seems immoral. Similarly, a college student might submit assignments on time to succeed academically but arrive early at class to socialize with other students. In both cases, a person who performed one of these behaviors and not the other would not necessarily be inconsistent, nor would a person who performed both necessarily be consistent. When ostensibly similar behaviors have different psychological determinants, perhaps consistency across them should not be expected in the first place.

Fortunately, there are other ways to assess behavior besides counting discrete and concrete acts. For example, trained coders have shown themselves to be quite adept at reliably coding behaviors in terms of their social effect or psychological meaning (Cairns & Green, 1979). This is an importantly different way of looking at things. It relocates the essence of behavior from its superficial appearance to the meaning it has for the person or the effect it has upon the social environment. It may be at this level of analysis that personality has its most important influence on action. What is also uncertain is whether behaviors assessed reliably but at such a higher level of psychological meaning and effect would manifest greater consistency than has been usually reported in the consistency literature. This issue will be the first specific concern of the present research.

Psychological Nature of Situations

Although it has frequently been observed that behavior is a function of the person and the situation and despite a rich tradition of developing technologies for assessing people (Wiggins, 1973), a long-standing lack in the field of personality assessment has been a well-developed technology for the assessment of situations (Bem & Funder, 1978).

This has been a key omission. The consistency controversy (Kenrick & Funder, 1988) that dominated personality psychology for more than two decades was specifically instigated by claims that behavior is determined more by properties of situations than of people (e.g., Mischel, 1968; Nisbett, 1980). Usually, the situationist case was made by subtraction: Behavioral variance that was not attributable to the person variables that were measured in a given study was assumed to be due to the situation instead (Hogan, DeSoto, & Solano, 1977). However, as Funder (1983) argued, "No science can progress far by accounting for a proportion of the variance with a variable you do measure, ascribing the much larger remainder to a variable you do not measure, and then asserting that the latter is more important than the former" (p. 288).

Bem and Funder (1978) suggested that one way to assess psychological properties of situations might be through their influence on behavior, as moderated by individual differences

in personality. Little research followed on their preliminary theoretical and empirical work, however.[1] A dozen years later, psychology still lacks a routine method for answering even this simple question: Which two of the three experimental situations described at the beginning of this article are psychologically the most similar to each other? The assessment of situations is the second specific concern of the present research.

Situational Specificity and Cross-Situational Consistency

Sometimes, psychologists have seemed to assume that behavioral consistency means people should behave in a constant fashion all the time. Taken to its logical conclusion, this point of view would imply that for a person to be consistent, the number of his or her smiles at a party and at a funeral must be the same! But, obviously, people change their behavior in important ways as situations change, and only someone who has severe psychological difficulties (e.g., a catatonic schizophrenic) will be consistent in the sense of being unchanging in his or her behavior regardless of circumstances (Mischel, 1968).

However, to conclude that personality variables are unimportant on the basis of observing how behavior changes across situations is to fall victim to a non sequitur (Lanning, 1988; Lay, 1977; Ozer, 1986). People might change their behavior markedly across situations and yet maintain a substantial degree of interindividual consistency, because cross-situational mean differences and cross-situational correlations are statistically independent. A specific, empirical demonstration of the relationship between cross-situational discriminativeness in behavior and cross-situational consistency is the third objective of the present research.

Differences Between Behaviors

A final concern is that behavioral consistency might itself be variable. This possibility has caused investigators to look for moderators of consistency. For instance, perhaps some people are more consistent than others (e.g., Bem & Allen, 1974; Chaplin & Goldberg, 1985; Zuckerman et al., 1988). The results of research exploring this possibility have been provocative but inconclusive. Another potential moderator is situational similarity: Perhaps behavior is consistent only across situations that are highly similar (Mischel & Peake, 1982). However, as mentioned earlier, the development of a technology to assess situational similarity has hardly begun (Bem & Funder, 1978).

A third potential moderator, which has received almost no attention, is the nature of behavior itself. Not all behaviors are the same. Different behaviors can result from vastly different causal mechanisms, be relevant to different personality dispositions, and perhaps even vary in the degree to which they are relevant to any personality disposition. This possibility will be the final topic of the present research.

Method

Subjects

The behavior of 140 undergraduate students, 70 of each sex, was observed and recorded on videotape in three laboratory situations.[2] All subjects were volunteers paid $4 per hour for their participation.

Three Laboratory Situations

Rationale. This research was designed to gather observations of each subject's behavior in three laboratory situations. This number is, of course, the minimum that will allow even a start toward assessing the properties of situations that might enhance or decrease cross-situational consistency. As George Kelly has stated, "in its minimum context a construct is a way in which two elements are similar and contrast with a third. There must therefore be at least three elements in the context" (G. A. Kelly, 1955, Vol. 1, p. 61). In the present context, at least three situations are required because that number begins to allow the situations across which behavior manifests greater and lesser consistency, to be compared for the physical and psychological properties ("constructs") they do and do not share.

The situations were designed to meet several criteria: (a) They had to allow the expression of individual differences in behavior (thus, they could not be overly controlling or "scripted"; cf. Schutte, Kenrick, & Sadalla, 1985; Snyder & Ickes, 1985), (b) they had to be procedurally simple and feasible within limited resources, (c) they could not deceive, and (d) the three situations had to differ from each other along identifiable, concrete, and potentially meaningful dimensions.[3] They were (a) an unstructured interaction with a member of the opposite sex (denoted Session 1), (b) a second unstructured interaction, a few weeks later, with a different member of the opposite sex (denoted Session 2), and (c) a contrived debate over capital punishment with the Session 2 partner (denoted debate). We used opposite-sex pairs to make the social interaction more involving for the participants and to balance the number of male and female subjects in all analyses.

The first two situations were similarly (un)structured but differed in two ways. First, the partners in the two sessions were different. Second, the subjects were more familiar with the research setting by Session 2 than they had been at Session 1. The second and third situations (Session 2 and debate) were similar in that subjects interacted with the same partner in both of them, and the sessions were held just a few minutes apart. The second situation was almost completely unstructured, however, whereas in the third situation, subjects were instructed to perform a specific task.[4]

The similarities and differences between these brief laboratory situations hardly exhaust the range and variety of the situations people encounter in real life. However, they provided an opportunity to observe directly and to videotape subjects as they performed actual social behavior. Moreover, as was noted earlier, even these three situations are sufficient to raise the question of which two are psychologically the most similar.

---

[1] Mischel and Peake (1982) reported a partial replication of the first of Bem and Funder's three studies, but their analyses and data were not relevant to Bem and Funder's broader thesis for reasons discussed by Funder (1983).

[2] A total of 164 subjects (plus informants) were recruited for this research over a period of 2½ years, making for an overall subject retention rate of 85%. Due to an error in coding assignments, the behavior of one male subject in the third situation was not coded.

[3] The project that gathered these data required an extraordinary amount of voluntary cooperation from our subjects over an extended period of time. Therefore, to deceive subjects at any point seemed both inappropriate and unwise.

[4] Subjects during Session 2 were not aware that the debate would follow immediately thereafter. The subjects' first and second experimental appointments were made for 1 hour, and each videotaped behavioral segment (one during the first appointment and two during the second) took only 5 min. The remaining time during the two appointments was occupied by questionnaire administration.

Session 1. From volunteer sign-up sheets, subjects were assigned randomly into opposite-sex pairs and contacted by telephone as to the time they should appear. When the second subject arrived for the laboratory session, the male experimenter escorted both subjects into a small room containing a couch, a video recorder, and a TV camera aimed at the couch (the camera was not concealed). The experimenter said:

> The first thing I would like you to do is just come on in here and get acquainted with each other. I would like to tape your conversation though. So just have a seat on the couch. Say whatever you like and I'll be back in about five minutes.

The experimenter then activated the video recorder in full view of the subjects and left the room. He returned 5 min later and turned off the equipment.

Session 2. Approximately 4 weeks later, subjects were again contacted and scheduled to appear in (different) opposite-sex pairs. As soon as the second subject arrived, both were shown into the laboratory room and the experiment proceeded exactly as before.

Debate. A few minutes after the conclusion of Session 2, the experimenter handed a clipboard to each subject and said:

> The next thing I do is hand each of you a pad of paper because some people like to be able to take some notes during the next part of the experiment. That is because the next part calls for the two of you to have a little debate. Specifically, the topic we have people debate is the use of capital punishment, because most people can come up with at least some arguments on both sides of that issue. I'll just flip a coin and have [name of subject] call it. If it comes up what you call, you will be in favor of capital punishment and if it doesn't then you will be against it. [The experimenter then flipped the coin.] Okay, the debate will last about another five minutes. I'll just give you a short minute to collect your thoughts and then we'll start.

After a brief pause, the experimenter said "begin," activated the video recorder, and left the room. He returned 5 min later and turned off the equipment.

Personality assessments. Each subject was asked to recruit two people who knew him or her well to come to the laboratory and complete personality descriptions of him or her. Specifically, the subject was instructed to seek the two people who knew him or her best among those currently in the vicinity and willing to participate. Two peers eventually appeared to describe each of 128 of the subjects who were videotaped in all three situations; 10 more of these subjects were described by a single peer. On the average, these "informants" had known their subject for 18.5 months. Of the informants, 56% described themselves as being primarily a friend of the subject, 33% as a roommate, 8% as a boyfriend/girlfriend, and 3% as other (e.g., sibling). All informants were paid for their participation and were assured (truthfully) that their description would not be made available to the acquaintance they described.

The personality descriptions were provided by completion of the California Q-Sort (Block, 1978), as slightly modified by Bem and Funder (1978) for use with nonprofessionals. The Q set consists of 100 descriptive statements about personality, each printed on a separate card. The judge's task is to sort the items into a forced, approximately normal, 9-category distribution that ranges from *not at all characteristic* (1) to *highly characteristic* (9) of the person judged. Analyses of interjudge agreement in the use of this instrument in this sample of subjects, both among peers' judgments and between peers' judgments and subjects' own self-judgments, are reported in articles by Funder and Dobroth (1987) and Funder and Colvin (1988). Analyses of correlations between these informants' judgments of personality and the subjects' scores on several measures of "social acuity" are reported by Funder and Harris (1986). Analyses of correlations between acquaintance and stranger judgments and measures of personality and behavior are reported by Colvin and Funder (in press).

Results

Behavioral Coding

The first step in data analysis was to code the behaviors exhibited in each laboratory situation into numerically analyzable form. This was a daunting and extremely time-consuming enterprise. The raw, unanalyzed stream of a person's behavior as recorded on videotape offers a virtually infinite number of coding possibilities, ranging from specific counts of small motor acts (eyebrow twitches, elbow lifts) to highly impressionistic judgments of the ultimate meaning of the subject's actions (Cairns & Green, 1979; Fiske, 1979). And, among the nearly limitless possibilities, no single coding scheme is the "right" one. Rather, the specific method needs to be determined by one's research goals (see Bakeman & Gottman, 1986, for a thorough and insightful discussion of these issues).

In the present research, our central goal was to extract units of behavior that would be psychologically meaningful in their own right and that would be relatively likely to be relevant to our subjects' general personality and behavior at large. At the same time, we needed a scheme that would require a minimum of subjective inference by coders and that could achieve a tolerable degree of reliability. These two goals oppose each other, of course. More mechanistic and objective coding schemes will be more reliable, but the coded behaviors will be relatively decontextualized and meaningless. More contextualized and subjective coding schemes will yield more meaningful behaviors, but reliability will necessarily suffer. Our task was to design a technique that would attain a reasonable compromise.

The coding scheme we finally developed took the form of a 62-item Q-sort deck (Block, 1978; Stephenson, 1953). The items were each written to describe categories of directly observable but meaningful behavior. Forty-one items describe behaviors directly relevant to characteristics included in the personality (California) Q set (CQ; Block, 1978). For example, one item in the CQ reads "is cheerful." The corresponding item written for the new, behavioral Q set (BQ) was "behaves in a cheerful manner." Another CQ item reads "has a readiness to feel guilt." The corresponding item in the BQ was, "expresses guilt (about anything)." Additional items were written to capture aspects of behavior that seemed important in the laboratory situations that were videotaped, even though they possibly were not directly relevant to items in the CQ (e.g., "expresses awareness of being on camera and/or in an experiment").

We had each coder watch (as many times as was necessary, at least twice) the 5-min videotape he or she was to code, and then as with other Q sorts, we had each coder arrange the cards of this deck into a forced, quasi-normal distribution ranging from *not at all or negatively characteristic of the behavior of the person in question* (1) to *highly characteristic of the person's behavior* (9). Coders were instructed to use the Q items to describe only behaviors they had witnessed on the videotape and to avoid, so far as was possible, "playing psychologist" or making inferences about subjects' general behavioral dispositions.

The numbers of cards in the 9 categories were, respectively, 3,

5, 7, 10, 12, 10, 7, 5, and 3. The forced sorting procedure automatically corrects for various rating response sets that coders might otherwise manifest (Block, 1978). Another important property of the Q sort is that it requires coders to judge the salience of each behavior in relation to all the others. Thus, the behaviors that receive the most extreme scores (1 or 9) are those judged to be the most important for characterizing how the individual acted in a particular videotaped situation, in relation to the other items in the set.

Each segment of videotaped behavior was coded independently by, on average, six coders. To keep estimates of cross-situational consistency unconfounded, no coder viewed more than a single behavioral episode for each subject. Partners within a single session were coded by different coders. The aggregate (Spearman-Brown) reliabilities of the BQ items ranged as high as .82, with a median reliability of .64. All but 4 items attained reliabilities of at least .40. A complete list of the 62 BQ items, arranged in order of their average estimated reliability across the three laboratory sessions, appears as Table 1.

Cross-Situational Consistency

Once the behaviors were coded, the next step was to ascertain whether our efforts to find a compromise between the reliability and meaningfulness of behavioral coding had paid off with higher cross-situational consistency coefficients than usual. The results appear in Tables 2, 3, and 4. (For purposes of comparison, the correlations as corrected for attenuation are included in these tables, as are the correlations calculated separately within each sex.)

As these tables reveal, we did find a substantial degree of cross-situational consistency in our subjects' behavior across the three laboratory situations. Between Sessions 1 and 2, a full 37, or more than half, of the 62 BQ items attained significant correlations at $p < .001$. The number of items reaching this level of consistency between Session 2 and the debate was 26; the number between Session 1 and the debate was 18.[5] Nearly all of these correlations were replicated independently within the female and male subsamples, as the tables attest.

Overall, 61 of the 62 cross-situational correlations between Sessions 1 and 2 were positive, and 45 were significant ($p < .05$). Between Session 2 and the debate, 57 correlations were positive, and 41 were significant ($p < .05$); between Session 1 and the debate, 56 correlations were positive, and 30 were significant ($p < .05$). None of the negative correlations approached significance. Across all 62 items, the three mean cross-situational correlations were, respectively, .37, .27, and .20.

The 62-item BQ has been shown to yield four factors, labeled Nervous Withdrawal, Domineeringness, Serious Intelligence, and Heterosexuality (Colvin & Funder, in press). All four behavioral factors, across all three cross-situational comparisons, yielded consistency coefficients that were significant with one exception. Between Sessions 1 and 2, the four consistency correlations were, respectively, .68, .53, .56, and .53. Between Session 2 and the debate, the four consistency correlations were, respectively, .55, .41, .42, and .40. Between Session 1 and the debate, the four consistency correlations were, respectively, .42, .20, .49, and .37 (the .20 just cited was the lone exception, being significant at the .05 level).

Some individual item correlations also were quite large. Independent codings of the behavior "exhibits an awkward interpersonal style" correlated .66 between Session 1 and Session 2, "laughs frequently" correlated .63, and "behaves in a cheerful manner" correlated .60. Between Session 2 and the debate, "behaves in a fearful or timid manner" correlated .60, and "has high enthusiasm and high energy level" correlated .53. Between Session 1 and the debate, "is expressive in face, voice, or gestures" correlated .52, and "is reserved and unexpressive" correlated .41.

These correlations are much greater than the highest consistency correlations reported in some of the most widely cited studies in the personality literature (e.g., Hartshorne & May, 1928; Mischel & Peake, 1982). Perhaps those studies looked for behavioral consistency in the wrong place. Instead of counting specific, microlevel behaviors, our coding scheme reliably assessed behaviors, such as fearfulness, in terms of their psychological meaning. What people manifest consistently across situations, the present results imply, may not be so much specific behaviors but underlying psychological dispositions that can be expressed behaviorally in numerous ways (cf. Cairns & Green, 1979).

Differences Between Situations

Despite the cross-situational consistency correlations just reported, which reflect how subjects maintained their individual differences in behavior across situations, our subjects might have been inconsistent in another sense, which is that they may have generally changed their behavior as a function of situational factors (Ozer, 1986). For instance, even though the same subjects who were the most fearful in one situation might also have been the most fearful in another, the second situation still might be, in general, much less fear evoking than the first.

Recall that the three laboratory situations in which the behavior of our subjects was videotaped were designed to be different from each other in three concrete ways. Sessions 1 and 2 were both unstructured interactions with opposite-sex strangers but involved different partners, and Session 2 occurred in a more familiar setting than did Session 1. Session 2 and the debate both involved interacting with the same partner, but whereas Session 2 was an unstructured interaction, the debate was structured as a contrived debate on capital punishment. Session 1 and the debate differed, of course, in all the concrete, procedural ways just mentioned.

One way to assess the psychological difference between these situations is simply to examine how the average behavior of our subjects varied across them. The behaviors that differed from each other at the .05 level or better between Sessions 1 and 2 are shown in Table 5, those that differed between Session 2 and the debate appear in Table 6, and those that differed between Session 1 and the debate appear in Table 7.

To highlight a couple of examples, subjects generally received higher scores on "behaves in a fearful manner" and "expresses

[5] All significance levels reported in this article are two-tailed values. A complete table of the correlations, means, and standard deviations for all 62 behavioral Q-sort items appears as an appendix to this article.

Table 1
Behavioral Q-Sort Items: Average and Individual Session Reliability Estimates

Behavioral Q-sort item	Average reliability	Session 1	Session 2	Session 3
16. High enthusiasm and high energy level.	.82	.82	.83	.81
10. Laughs frequently (whether "genuine" or "nervous").	.82	.79	.79	.87
9. Is reserved and unexpressive.	.81	.80	.81	.82
57. Speaks in a loud voice.	.79	.76	.82	.80
6. Dominates the interaction.	.78	.80	.73	.80
37. Behaves in a fearful or timid manner.	.77	.78	.77	.77
11. Smiles frequently.	.76	.72	.76	.80
30. Seeks advice from partner (low = *partner seeks advice from subject*)	.75	.81	.85	.51
14. Exhibits an awkward interpersonal style.	.75	.78	.72	.74
50. Behaves in a cheerful manner.	.75	.76	.73	.76
60. Engages in constant eye contact with partner.	.75	.75	.72	.78
61. Shows lack of interest in the interaction.	.74	.73	.75	.74
21. Is talkative (in this situation).	.73	.74	.66	.77
1. Expresses awareness of being on camera/in experiment.	.71	.78	.74	.58
26. Initiates humor.	.71	.63	.73	.76
62. Speaks quickly.	.70	.68	.66	.76
38. Is expressive in face, voice, or gestures.	.70	.65	.70	.74
12. Physically animated; moves around a great deal.	.69	.69	.70	.67
19. Expresses agreement unusually frequently.	.69	.59	.66	.79
2. "Interviews" partner (e.g., asks series of questions.)	.68	.67	.73	.65
23. Shows physical signs of tension or anxiety.	.67	.67	.63	.70
7. Appears to be relaxed and comfortable.	.66	.70	.65	.62
42. Shows genuine interest in intellectual matters.	.66	.70	.63	.65
41. Keeps partner at a distance.	.66	.67	.69	.61
52. Behaves in a masculine or feminine style or manner.	.66	.66	.67	.65
24. Exhibits high degree of intelligence.	.66	.66	.60	.70
53. Offers advice to partner.	.65	.77	.81	.20
22. Expresses insecurity or sensitivity.	.65	.68	.62	.66
43. Seems genuinely to enjoy interaction with partner.	.64	.65	.65	.62
3. Volunteers unusually little information about self.	.64	.64	.66	.62
5. Tries to control the interaction.	.63	.59	.55	.73
59. Makes physical contact with partner.	.62	.70	.64	.50
8. Exhibits social skills.	.62	.68	.64	.54
28. Exhibits condescending behavior.	.62	.60	.60	.66
4. Seems genuinely interested in what partner has to say.	.61	.63	.66	.54
13. Seems to genuinely like the partner.	.60	.59	.59	.61
54. Speaks fluently and expresses ideas well.	.59	.59	.48	.69
36. Is unusual or unconventional in appearance.	.58	.59	.64	.49
35. Expresses hostility.	.58	.55	.52	.65
29. Partner seems to like him or her.	.58	.54	.58	.61
31. Appears to regard self physically attractive.	.57	.55	.65	.49
15. Interrupts partner (low = *partner interrupts subject*)	.55	.57	.32	.71
33. Expresses warmth	.55	.56	.54	.56
18. Talks at rather than with partner (e.g., *monologue*)	.54	.51	.54	.58

Table 1 (*continued*)

Behavioral Q-sort item	Average reliability	Session 1	Session 2	Session 3
20. Expresses skepticism or cynicism.	.53	.59	.61	.38
56. Demonstrates interest in competition.	.53	.66	.63	.22
51. Discusses philosophical issues with interest.	.52	.57	.38	.58
32. Acts in an irritable fashion.	.50	.47	.45	.58
55. Brags.	.49	.60	.54	.29
25. Expresses sympathy toward partner.	.48	.56	.34	.51
48. Expresses self-pity or feelings of victimization.	.48	.42	.60	.41
40. Expresses guilt (about anything).	.47	.40	.53	.48
46. Displays ambition.	.46	.69	.56	.03
58. Demonstrates interest in topics related to power.	.43	.52	.37	.40
17. Discusses unusually large number of topics.	.43	.40	.54	.33
27. Seeks reassurance from partner.	.42	.55	.35	.33
34. Tries to sabotage or obstruct experiment or partner.	.42	.37	.49	.39
45. Shows interest in conventional ways of judging people.	.40	.41	.39	.41
39. Expresses interest in fantasy and daydreams.	.35	.47	.35	.22
44. Says or does interesting things (from partner's point of view)	.35	.40	.41	.24
49. Seems interested in partner as member of opposite sex.	.29	.48	.36	.01
47. Seems to view interaction as sexual encounter.	.14	.28	.12	.00

Note. Items are arranged in order of their average estimated reliability across the three laboratory sessions.

insecurity or sensitivity" in Session 1, which was a unique context for them, compared with Session 2, by which time they were more experienced and, it would seem from Table 5, more comfortable. Consistent with this interpretation, the behaviors coded as "exhibits social skills" and "behaves in a cheerful manner" were rated significantly more highly in Session 2 than they were in Session 1. In comparisons between Session 2 (an unstructured interaction) and the debate, subjects were rated higher in Session 2 on "expresses warmth," "seems to like partner," and "initiates humor," whereas their behavior in the debate was more likely to involve "discuss[ing] philosophical issues with interest," "express[ing] skepticism or cynicism," and "act[ing] in an irritable fashion"—findings that reflect the different goals subjects pursued in the debate, compared with the friendly interactions of Session 2. The differences between Session 1 and the debate were, as might be expected, essentially a compound of the differences between Sessions 1 and 2 as well as Session 2 and the debate.

These cross-situational mean differences suggest a tentative answer to the question raised earlier, which was, which two of these three situations are psychologically the most similar to each other, and which two are the most different? Apparently, Session 1 was more psychologically similar to Session 2 than Session 2 was to the debate. Of the 62 behaviors coded, fully 40 manifested significant differences between Session 2 and the debate, and 38 differed between Session 1 and the debate, whereas about half that number (a still substantial 20) were significantly different between Sessions 1 and 2. According to the test for correlated proportions (see Funder & Colvin, 1988; Guilford & Fruchter, 1978; McNemar, 1947), significantly more differences between Session 1 and the debate and between Session 2 and the debate were significant than between Session 1 and Session 2 ($zs = 3.25$ and 2.72, respectively, $ps < .001$).[6]

The mean differences in behavior also demonstrate two more general points. First, even though an impressive degree of cross-situational consistency was found across the three situations, they differed from each other both operationally and psychologically. Second, mean differences and cross-situational correlations are essentially uncorrelated. At the item level, the average cross-situational correlation of a given BQ item across the three situations correlated with its average mean change (calculated in terms of absolute values), $r = -.02$, *ns.*

Differences Between Behaviors

Consistency across laboratory situations. Yet another finding emerged when we looked again at the cross-situational consis-

[6] This conclusion is congruent with the findings, reported in the previous section, that more cross-situational consistency correlations were significant between Session 1 and Session 2 than between Session 2 and the debate or between Session 1 and the debate. These differences, too, were all significant at the .001 level, according to the test for correlated proportions (between Session 1–Session 2 and Session 2–debate, $z = 2.67$; between Session 1–Session 2 and Session 1–debate, $z = 4.69$; and between Session 2–debate and Session 1–debate, $z = 3.05$).

Table 2
Cross-Situational Consistency Correlations Across Sessions 1 and 2 for Total Sample and by Sex of Subject

Behavioral Q-sort item	Total sample	Total sample (disatten.)	Women	Men
57. Speaks in a loud voice.	.70	.89	.74	.67
14. Exhibits an awkward interpersonal style.	.60	.88	.67	.62
37. Behaves in a fearful or timid manner.	.65	.84	.68	.62
10. Laughs frequently (whether genuine or nervous).	.63	.80	.56	.63
38. Is expressive in face, voice, or gestures.	.63	.93	.65	.58
9. Is reserved and unexpressive.	.62	.77	.64	.57
11. Smiles frequently.	.60	.81	.39	.60
50. Behaves in a cheerful manner.	.60	.81	.52	.60
16. High enthusiasm and high energy level.	.59	.72	.55	.59
62. Speaks quickly.	.59	.88	.55	.60
8. Exhibits social skills.	.58	.88	.56	.59
60. Engages in constant eye contact with partner.	.57	.78	.59	.53
22. Expresses insecurity or sensitivity.	.56	.86	.52	.60
31. Appears to regard self physically attractive.	.55	.92	.55	.53
61. Shows lack of interest in the interaction.	.54	.73	.44	.62
7. Appears to be relaxed and comfortable.	.48	.71	.48	.48
28. Exhibits condescending behavior.	.47	.78	.56	.40
23. Shows physical signs of tension or anxiety.	.45	.69	.40	.50
36. Is unusual or unconventional in appearance.	.45	.73	.29	.55
24. Exhibits high degree of intelligence.	.44	.70	.39	.48
32. Acts in an irritable fashion.	.43	.94	.35	.48
52. Behaves in a masculine or feminine style or manner.	.43	.65	.39	.45
43. Seems genuinely to enjoy interaction with partner.	.42	.65	.27	.53
54. Speaks fluently and expresses ideas well.	.42	.79	.40	.41
26. Initiates humor.	.41	.60	.40	.40
20. Expresses skepticism or cynicism.	.40	.67	.34	.43
12. Physically animated; moves around a great deal.	.39	.56	.34	.43
41. Keeps partner at a distance.	.39	.57	.29	.48
5. Tries to control the interaction.	.38	.67	.41	.33
18. Talks at rather than with partner (e.g., monologue).	.38	.72	.24	.48
19. Expresses agreement unusually frequently.	.38	.61	.40	.35
21. Is talkative (in this situation).	.38	.54	.46	.32
42. Shows genuine interest in intellectual matters.	.36	.54	.22	.46
4. Seems genuinely interested in what partner has to say.	.34	.53	.27	.40
35. Expresses hostility.	.30	.56	.31	.25
53. Offers advice to partner.	.29	.37	.41	.20
48. Expresses self-pity or feelings of victimization.	.28	.56	.37	.22

Note. $N = 140$. Items with consistency correlations of $p < .001$ (two-tailed) or better are listed. Items are arranged in order of their cross-situational consistency. Disatten. = disattenuated.

tency correlations: The same behaviors that were highly consistent across one pair of situations tended to be highly consistent across the other two pairs as well. Specifically, the cross-situational consistency correlations between Sessions 1 and 2 correlated .73 with those calculated between Session 2 and the debate and correlated .75 with those between Session 1 and the debate, and the cross-situational consistency of behavioral items between Session 2 and the debate correlated .84 (!) with the consistency of items between Session 1 and the debate. The consistency of behavior, it would seem, is itself highly consis-

Table 3
Cross-Situational Consistency Correlations Across Session 2 and Debate for Total Sample and by Sex of Subject

Behavioral Q-sort item	Total sample	Total sample (disatten.)	Women	Men
57. Speaks in a loud voice.	.65	.80	.69	.61
36. Is unusual or unconventional in appearance.	.61	.99	.45	.69
37. Behaves in a fearful or timid manner.	.60	.78	.67	.54
60. Engages in constant eye contact with partner.	.60	.80	.56	.67
52. Behaves in a masculine or feminine style or manner.	.54	.82	.39	.67
16. High enthusiasm and high energy level.	.53	.65	.62	.43
38. Is expressive in face, voice, or gestures.	.53	.74	.56	.50
62. Speaks quickly.	.53	.75	.60	.55
9. Is reserved and unexpressive.	.52	.64	.50	.52
12. Physically animated; moves around a great deal.	.47	.69	.19	.61
23. Shows physical signs of tension or anxiety.	.46	.69	.52	.39
7. Appears to be relaxed and comfortable.	.45	.71	.39	.52
10. Laughs frequently (whether genuine or nervous).	.43	.52	.33	.46
31. Appears to regard self physically attractive.	.42	.74	.45	.39
50. Behaves in a cheerful manner	.41	.55	.34	.43
43. Seems genuinely to enjoy interaction with partner.	.40	.63	.53	.30
11. Smiles frequently.	.39	.50	.33	.40
22. Expresses insecurity or sensitivity.	.38	.59	.51	.24
29. Partner seems to like him or her.	.34	.57	.49	.23
6. Dominates the interaction.	.32	.42	.38	.24
61. Shows lack of interest in the interaction.	.32	.43	.48	.20
14. Exhibits an awkward interpersonal style.	.31	.42	.45	.21
41. Keeps partner at a distance.	.31	.48	.29	.32
32. Acts in an irritable fashion.	.30	.59	.36	.27
13. Seems to genuinely like the partner.	.29	.48	.26	.30
24. Exhibits high degree of intelligence.	.29	.45	.15	.40

Note. $N = 139$. Items with consistency correlations of $p < .001$ (two-tailed) or better are listed. Items are arranged in order of their cross-situational consistency. Disatten. = disattenuated.

tent. The 15 most and least consistent behavioral items, averaged across the three cross-situational analyses, appear in Table 8.

Consistency between laboratory settings and personality judgments. To check against the possibility that it was merely some peculiarity of the three laboratory situations that led some behaviors to be more consistent than others, we compared the consistency of our subjects' behavior across our three laboratory situations with the consistency of their behavior between the laboratory situations and real life.

Recall that 41 of the BQ items were written to have close analogues in the CQ with which subjects were described by their acquaintances. This fact allows us to correlate each of these 41 BQ items, derived from behavior in the lab, with each of the corresponding 41 CQ items, derived by means of acquaintances' judgments that are based on behaviors observed in real life. The same behavioral items that attained high correlations across our three laboratory situations also tended to attain high correlations between the laboratory and real life (relevant $r = .35$, $p < .03$) and vice versa. The 15 highest and 15 lowest correlations appear in Table 9.

Many of these correlations are significant, although they tend to be smaller than the cross-situational correlations cited earlier. The higher correlations can be taken to imply that the behavior we observed in our three laboratory situations, even though it occurred in a setting outside of the mainstream of daily life, was not entirely artificial. Subjects' behavior within the laboratory was meaningfully related to their behavior outside. For instance, the degree to which subjects showed "genuine interest in intellectual matters," as measured in the labora-

Table 4
Cross-Situational Consistency Correlations Across Session 1 and Debate for Total Sample and by Sex of Subject

Behavioral Q-sort item	Total sample	Total sample (disatten.)	Women	Men
57. Speaks in a loud voice.	.58	.74	.55	.61
62. Speaks quickly.	.56	.78	.45	.59
38. Is expressive in face, voice, or gestures.	.52	.75	.42	.58
52. Behaves in a masculine or feminine style or manner.	.48	.73	.48	.48
16. High enthusiasm and high energy level.	.47	.58	.50	.41
36. Is unusual or unconventional in appearance.	.47	.87	.38	.52
37. Behaves in a fearful or timid manner.	.45	.58	.52	.38
31. Appears to regard self physically attractive.	.43	.83	.45	.41
60. Engages in constant eye contact with partner.	.43	.56	.40	.45
9. Is reserved and unexpressive.	.41	.51	.39	.41
12. Physically animated; moves around a great deal.	.38	.56	.28	.43
24. Exhibits high degree of intelligence.	.34	.50	.29	.38
42. Shows genuine interest in intellectual matters.	.32	.47	.18	.47
11. Smiles frequently.	.30	.40	.29	.25
22. Expresses insecurity or sensitivity.	.30	.45	.32	.28
14. Exhibits an awkward interpersonal style.	.28	.37	.35	.23
23. Shows physical signs of tension or anxiety.	.28	.41	.26	.30
50. Behaves in a cheerful manner.	.28	.37	.24	.26

Note. $N = 139$. Items with consistency or correlations of $p < .001$ (two-tailed) or better are listed. Items are arranged in order of their cross-situational consistency. Disatten. = disattenuated.

tory, correlated .40 ($p < .001$) with the same behavioral style as assessed by acquaintances from daily life. Being "reserved and unexpressive" in the laboratory correlated .32 ($p < .001$) with the same behavioral style as assessed by acquaintances. And so forth.

Behavior certainly was more consistent among the three lab situations, however, than it was between the lab and real life. Even though we have documented that the three laboratory situations were substantially different from each other operationally and psychologically, it seems quite obvious that the psychological distance between the lab situations and the real-life settings that were the basis of the informants' ratings still was greater than was this distance among the lab situations themselves. Accordingly, it is only natural that consistency would be less. Moreover, the people who coded behavior from the videotapes all viewed the same 5-min segment, which they could watch as many times as necessary to record it accurately. The friends and acquaintances, by contrast, based their assessments on their diverse and idiosyncratic experience with the subject and relied solely on memory. Also, the behavioral coders received extensive training in the behavioral manifestations by which they were to code each item in the BQ. Friends and acquaintances were allowed to use their own, possibly idiosyncratic, definitions of each CQ item. These procedural differences probably also contributed to the differences between the two kinds of cross-situational correlation.

Explanations of the Difference Between Behaviors

Psychometric explanation. According to the preceding analyses, some behaviors seem to be more consistent than others, both across the three lab situations and between the lab and real life. But an elementary principle of psychometrics is that correlations (including consistency correlations) will be constrained by the reliabilities and variances of the variables they correlate (Wiggins, 1973). Hence, one possible explanation for the difference between behaviors we have uncovered is simply that some behaviors are coded more reliably, or exhibit wider variance across subjects, than do other behaviors and so are able to manifest larger cross-situational correlations.

Support certainly can be found for this point of view. The average variance of the 62 BQ items (that is, the average of the three between-subjects variances calculated within each of the three laboratory situations) correlates .75 ($p < .001$) with the average consistency of these items across the three situations and .33 ($p < .05$) with their consistency between lab settings and general personality judgments. A plausible and psychometrically sound possibility, therefore, is that BQ items with

Table 5
Significant Behavioral Mean Differences Between Sessions 1 and 2

Behavioral Q-sort item	Session 1 *M*	Session 2 *M*	*t*
Higher Session 1 means			
18. Talks at rather than with partner (e.g., monologue).	3.98	3.51	4.96***
14. Exhibits an awkward interpersonal style.	4.19	3.60	4.50
23. Shows physical signs of tension or anxiety.	5.19	4.66	3.76
61. Shows lack of interest in the interaction.	3.98	3.55	3.33
41. Keeps partner at a distance.	4.81	4.40	2.97**
22. Expresses insecurity or sensitivity.	4.77	4.49	2.93
37. Behaves in a fearful or timid manner.	3.98	3.64	2.85
24. Exhibits high degree of intelligence.	5.39	5.24	2.24*
12. Physically animated; moves around a great deal.	3.85	3.56	2.14
32. Acts in an irritable fashion.	3.76	3.60	1.95
Higher Session 2 means			
8. Exhibits social skills.	5.94	6.46	−4.65***
7. Appears to be relaxed and comfortable.	5.56	6.13	−3.98
44. Says or does interesting things (from partner's point of view).	5.78	6.08	−2.79**
38. Is expressive in face, voice, or gestures.	5.11	5.42	−2.68
2. Interviews partner (e.g., asks series of questions)	5.83	6.21	−2.56*
54. Speaks fluently and expresses ideas well.	5.98	6.25	−2.38
21. Is talkative (in this situation).	5.73	6.05	−2.33
60. Engages in constant eye contact with partner.	6.08	6.37	−2.32
43. Seems genuinely to enjoy interaction with partner.	5.90	6.14	−2.00
50. Behaves in a cheerful manner.	5.89	6.11	−1.94

Note. $N = 140$, $df = 138$. All tests were two-tailed. Items are arranged according to the significance of their mean difference across sessions.
* $p < .05$ for absolute value of $t = 1.94$ to 2.56. ** $p < .01$ for absolute value of $t = 2.68$ to 2.97. *** $p < .001$ for absolute value of $t = 3.33$ to 4.96.

larger within-session variances yielded greater reliabilities, as would be expected on purely statistical grounds, and therefore also manifested higher correlations across lab situations and between lab situations and general personality judgments reflective of behavior in real life.

However, this psychometric explanation is not a sufficient explanation for the difference between behaviors. If we partial both item variance and reliability from the correlations that index the stability of the difference between behaviors across the three laboratory situations, the correlations that were .73, .75, and .84 remain the still highly significant .40, .51, and .71, respectively. Similarly, the correlation between the consistency of average behaviors across the three situations and their consistency with personality as measured by the CQ changes only a little, from .35 to .29, when both variability and reliability are partialed from the BQ items.

Breadth of situational relevance. A second perusal of the items in Tables 8 and 9 reveals that the more consistent items seem to describe behaviors that are observed more often than the less consistent items. That is, items such as "speaks in a loud voice" and "behaves in a fearful or timid manner" describe behaviors that are relevant to a broader range of situations in real life than items such as "expresses interest in fantasy and daydreams" or "demonstrates interest in topics related to power."

To check this hypothesis, we obtained ratings of each of the 62 BQ items in response to the question, "in how many situations can each behavior occur?" To provide their answers, a group of six raters, working independently, sorted the 62 BQ items into a forced, nearly rectangular, 9-step distribution ranging from *very few situations in real life* to *most or all situations in real life.* The numbers of items in the 9 categories were, respectively, 7, 7, 7, 7, 6, 7, 7, 7, and 7. The aggregate reliability of the ratings was .79.[7]

[7] This item property is similar in some ways to the trait visibility property found to affect interjudge agreement in earlier studies by Funder and Dobroth (1987) and Funder and Colvin (1988): Visibility correlates with commonness .67 ($p < .001$). However, the properties are distinctive as well. Visibility does not perform nearly as well as commonness as a moderator of cross-situational consistency. Moreover, none of the correlations to be reported later in this article, concerning moderators of consistency, change in any substantial way when the influence of visibility is partialed from them.

Table 6
Significant Behavioral Mean Differences Between Session 2 and Debate

Behavioral Q-sort item	Session 2 *M*	Debate *M*	*t*
Higher Session 2 means			
25. Expresses sympathy toward partner.	5.34	4.71	8.46***
2. "Interviews" partner (e.g. asks series of questions).	6.23	5.04	7.23
33. Expresses warmth.	5.44	4.84	6.66
46. Displays ambition.	5.34	4.80	6.43
48. Expresses self-pity or feelings of victimization.	4.82	4.24	6.40
29. Partner seems to like him or her.	6.21	5.61	6.16
56. Demonstrates interest in competition.	5.47	4.97	6.10
13. Seems to genuinely like the partner.	6.01	5.45	5.66
19. Expresses agreement unusually frequently.	5.43	4.72	5.19
43. Seems genuinely to enjoy interaction with partner.	6.13	5.62	4.20
26. Initiates humor.	5.37	4.88	3.90
1. Expresses awareness of being on camera/in experiment.	5.35	4.87	3.38
11. Smiles frequently.	6.14	5.68	3.38
49. Seems interested in partner as member of opposite sex.	4.44	4.20	3.23**
55. Brags.	4.91	4.66	3.17
59. Makes physical contact with partner.	4.49	4.17	2.94
50. Behaves in a cheerful manner.	6.09	5.69	2.93
40. Expresses guilt (about anything).	4.48	4.30	2.28*
37. Behaves in a fearful or timid manner.	3.65	3.39	2.13
39. Expresses interest in fantasy and daydreams.	4.97	4.90	2.13
47. Seems to view interaction as sexual encounter.	4.24	4.11	1.96
Higher debate means			
51. Discusses philosophical issues with interest.	4.93	6.18	−11.77***
58. Demonstrates interest in topics related to power.	4.92	5.50	−8.00
20. Expresses skepticism or cynicism.	4.84	5.58	−6.76
57. Speaks in a loud voice.	4.28	4.94	−5.98
42. Shows genuine interest in intellectual matters.	5.44	5.89	−4.25
32. Acts in an irritable fashion.	3.60	4.00	−3.83
38. Is expressive in face, voice, or gestures.	5.40	5.92	−3.83
41. Keeps partner at a distance.	4.41	4.92	−3.74
5. Tries to control the interaction.	4.43	4.94	−3.50
12. Physically animated; moves around a great deal.	3.55	3.97	−3.49
45. Shows interest in conventional ways of judging people.	5.25	5.54	−3.37
18. Talks at rather than with partner (e.g., monologue).	3.52	3.93	−3.36
3. Volunteers unusually little information about self.	4.56	5.01	−3.32
28. Exhibits condescending behavior.	4.65	5.03	−3.31
24. Exhibits high degree of intelligence.	5.24	5.58	−3.29
17. Discusses unusually large number of topics.	4.51	4.86	−2.69**

Table 6 (*continued*)

Behavioral Q-sort item	Session 2 *M*	Debate *M*	*t*
35. Expresses hostility.	4.25	4.51	−2.31*
16. High enthusiasm and high energy level.	4.58	4.90	−2.05
21. Is talkative (in this situation).	6.05	6.35	−1.99

Note. $N = 139$, $df = 138$. All tests were two-tailed. Items are arranged according to the significance of their mean difference across sessions.
* $p < .05$ for absolute value of $t = 1.96$ to 2.31. ** $p < .01$ for absolute value of $t = 2.69$ to 3.23. *** $p < .001$ for absolute value of $t = 3.29$ to 11.77.

Across all 62 BQ items, breadth ratings correlated .50 ($p < .001$) with the items' variability and .53 ($p < .001$) with the items' reliability. The breadth ratings also correlated .55 ($p < .001$) with the average consistency of the 62 items across the three laboratory situations, and among the 41 BQ items that had direct CQ analogues, breadth correlated .37 ($p < .05$) with the consistency of these items between the laboratory situations and behavior in daily life.

Thus, one property of the content of our BQ items that underlies item variability, reliability, and cross-situational consistency apparently is the breadth of the range of situations to which they are relevant. A further pair of questions follows close on the heels of this finding: Why should more broadly relevant behaviors be more variable? And what does the breadth of a behavior's relevance have to do with its cross-situational consistency?

Operants versus respondents? A few years ago, David McClelland proposed a possible difference between behaviors that he would and would not expect to be consistent across situations. In an intriguing article entitled "Is Personality Consistent?" McClelland (1984), borrowing a couple of terms from Skinner (1931, 1938/1966), distinguished between "respondents . . . [defined as] responses to clearly identified stimuli" and "operants . . . [defined as] thoughts or actions the stimulus for which cannot be readily identified" (p. 194). McClelland pointed out that by these definitions, one would expect operant behaviors to express aspects of personality that are generally influential across diverse situations but expect respondent behaviors to be more responsive to the exact stimuli present in each setting. And therefore, he argued, it would be through operant behaviors that cross-situational consistency, and personality itself, would be more clearly manifest.

Over the years since Skinner (1931) proposed the distinction between respondents and operants, a great deal of excess baggage has, to be sure, attached itself to these terms. For instance, as the distinction developed into what became called "two-factor theory," respondents took on connotations of being behaviors that were innate, unconditioned, and physiologically based, and operants became viewed as being behaviors that were under more "voluntary" control (Bower & Hilgard, 1981, p. 200). For instance, Teitelbaum (1977) claimed that a respondent "by definition. . . excludes motivation. It is unconscious, unlearned, and involuntary" (p. 13).

The original distinction carried none of these implications, however. Skinner went out of his way to specifically decry "the unfortunate historical definition of the reflex [respondent] as a form of movement unconscious, involuntary and unlearned" (1931, p. 455). His preferred definition was more simple: "The kind of behavior that is correlated with specific eliciting stimuli may be called *respondent* behavior and a given correlation *a respondent*. The term is intended to carry the sense of a relation to a prior event." About operants, the other class of behavior, Skinner said, "An operant is an identifiable part of behavior of which it may be said, not that no stimulus can be found that will elicit it, but that no correlated stimulus can be detected upon occasions when it is observed to occur" (1938/1966, p. 21); "the original 'spontaneous' activity of the organism is chiefly of this sort, as is the greater part of the. . . behavior of the adult organism" (1938/1966, p. 19).

In that last sentence, Skinner implies that operants occur across a wider range of situations than do respondents. McClelland claims that operants are more consistent than respondents. Could this be the basis of a relationship between behavioral variability, breadth of situational relevance, and cross-situational consistency?

As a first effort toward answering this question, we asked nine raters to evaluate the degree to which each of the 62 BQ items described an operant as opposed to respondent behavior.[8] We did not use these terms, however. Rather, we asked our judges to evaluate each item in relation to what Skinner (and McClelland) presented as the essence of the distinction: The degree to which each behavior tended to occur in response to a specific, identifiable stimulus. The raters sorted the 62 items into a nearly rectangular Q distribution ranging from *stimulus bound, i.e., having a clearly identifiable stimulus* (1) to *stimulus free, i.e., are a characteristic style that people possess* (9). The aggregate reliability of their ratings was .86. Three items that received particularly high operant ratings were "speaks fluently," "is expressive in face, voice, or gestures," and "is unusual or unconventional in appearance." Three relatively respondent items were "seems interested in partner as a member of the opposite sex," "seems genuinely interested in what the partner has to say," and "expresses sympathy toward partner."

The final step was to integrate these ratings with our analyses

[8] In defense of rating the distinction along a continuous scale, we may quote Skinner himself to the effect that "although a distinction may be drawn between the operant and the respondent field, there is also a certain continuity" (1938/1966, p. 439).

Table 7
Significant Behavioral Mean Differences Between Session 1 and Debate

Behavioral Q-sort item	Session 1 *M*	Debate *M*	*t*
Higher Session 1 means			
46. Displays ambition.	5.55	4.80	7.04***
25. Expresses sympathy toward partner.	5.26	4.71	6.79
48. Expresses self-pity or feelings of victimization.	4.70	4.24	5.51
19. Expresses agreement unusually frequently.	5.46	4.72	5.35
33. Expresses warmth.	5.32	4.84	5.29
1. Expresses awareness of being on camera/in experiment.	5.69	4.87	5.09
29. Partner seems to like him or her.	6.15	5.61	5.03
2. "Interviews" partner (e.g., asks series of questions).	5.82	5.04	4.93
56. Demonstrates interest in competition.	5.33	4.97	4.13
37. Behaves in a fearful or timid manner.	3.99	3.39	4.05
22. Expresses insecurity or sensitivity.	4.78	4.29	3.89
40. Expresses guilt (about anything).	4.55	4.30	3.52
13. Seems to genuinely like the partner.	5.82	5.45	3.26
26. Initiates humor.	5.27	4.88	3.03**
61. Shows lack of interest in the interaction.	3.99	3.46	3.02
55. Brags.	4.88	4.66	2.67
9. Is reserved and unexpressive.	4.43	3.92	2.64
59. Makes physical contact with partner.	4.45	4.17	2.36*
11. Smiles frequently.	6.02	5.68	2.30
10. Laughs frequently (whether genuine or nervous).	5.73	5.36	2.27
23. Shows physical signs of tension or anxiety.	5.22	4.85	2.19
Higher debate means			
51. Discusses philosophical issues with interest.	4.99	6.18	−10.94***
58. Demonstrates interest in topics related to power.	5.00	5.50	−7.14
57. Speaks in a loud voice.	4.12	4.94	−7.10
38. Is expressive in face, voice, or gestures.	5.10	5.92	−5.97
20. Expresses skepticism or cynicism.	4.97	5.58	−5.56
7. Appears to be relaxed and comfortable.	5.53	6.38	−4.95
45. Shows interest in conventional ways of judging people.	5.16	5.54	−4.31
3. Volunteers unusually little information about self.	4.44	5.01	−4.10
5. Tries to control the interaction.	4.35	4.94	−3.63
21. Is talkative (in this situation).	5.73	6.35	−3.61
42. Shows genuine interest in intellectual matters.	5.61	5.89	−2.66**
62. Speaks quickly.	4.85	5.12	−2.64
16. High enthusiasm and high energy level.	4.52	4.90	−2.35*
8. Exhibits social skills.	5.93	6.27	−2.29
28. Exhibits condescending behavior.	4.79	5.03	−2.06
35. Expresses hostility.	4.27	4.51	−1.98
32. Acts in an irritable fashion.	3.77	4.00	−1.94

Note. $N = 139$, $df = 138$. All tests were two-tailed. Items are arranged according to the significance of their mean difference across sessions.
* $p < .05$ for absolute value of $t = 1.94$ to 2.36. ** $p < .01$ for absolute value of $t = 2.64$ to 3.03. *** $p < .001$ for absolute value of $t = 3.26$ to 10.94.

Table 8
Most and Least Consistent Behavioral Q-Sort Items, Averaged Across 3 Situational Comparisons

Behavioral Q-sort item	Average *r*	Operant/respondent score
15 most consistent		
57. Speaks in a loud voice.	.65	67
37. Behaves in a fearful or timid manner.	.57	41
38. Is expressive in face, voice, or gestures.	.56	74
62. Speaks quickly.	.56	67
60. Engages in constant eye contact with partner.	.54	38
16. High enthusiasm and high energy level.	.53	67
9. Is reserved and unexpressive.	.52	56
36. Is unusual or unconventional in appearance.	.51	72
52. Behaves in a masculine or feminine style or manner.	.48	65
31. Appears to regard self physically attractive.	.47	56
10. Laughs frequently (whether "genuine" or "nervous").	.46	44
50. Behaves in a cheerful manner.	.44	66
11. Smiles frequently.	.44	50
14. Exhibits an awkward interpersonal style.	.44	62
22. Expresses insecurity or sensitivity.	.42	46
M	.51	58.07
15 least consistent		
39. Expresses interest in fantasy and daydreams.	−.01	50
2. "Interviews" partner (e.g., asks series of questions).	.01	45
49. Seems interested in partner as member of opposite sex.	.04	14
47. Seems to view interaction as sexual encounter.	.05	17
34. Tries to sabotage or obstruct experiment or partner.	.06	44
33. Expresses warmth.	.06	49
58. Demonstrates interest in topics related to power.	.08	43
55. Brags.	.08	47
51. Discusses philosophical issues with interest.	.08	51
17. Discusses unusually large number of topics.	.09	35
3. Volunteers unusually little information about self.	.10	43
40. Expresses guilt (about anything).	.10	41
46. Displays ambition.	.11	55
30. Seeks advice from partner (low = *partner seeks advice from subject*)	.11	27
15. Interrupts partner (low = *partner interrupts subject*)	.12	38
M	.07	39.93

of cross-situational consistency. The results were rather striking. Higher scores reflected a rating of the behavior as being more operant, and lower scores reflected a rating of the behavior as being more respondent, as we have defined the terms. The 15 most consistent behaviors across the three laboratory situations, as listed in Table 8, were rated significantly higher than the 15 least consistent behaviors, $t(28) = 4.17$, $p < .001$. Similarly, the 15 BQ items that manifested the most consistency with matching personality items as judged from experience outside the lab, listed in Table 9, received higher scores than the 15 items with the least consistency between the lab and real life, $t(28) = 4.56$, $p < .001$.

Table 9
Highest and Lowest Personality/Behavior Correlations, With Operant/Respondent Scores

Behavioral Q-sort item	Personality Q-sort item	*r*	Operant/respondent score
	15 highest items		
31. Appears to regard self physically attractive.	31. Regards self as physically attractive	.41	56
42. Shows genuine interest in intellectual matters.	51. Values intellectual and cognitive matters	.40	64
24. Exhibits high degree of intelligence.	8. High intellect	.34	68
46. Displays ambition.	71. High aspiration level	.33	55
9. Is reserved and unexpressive.	97. Emotionally bland	.32	56
52. Behaves in a masculine or feminine style or manner.	93. Sex-typed	.32	65
8. Exhibits social skills.	92. Has social poise	.31	63
50. Behaves in a cheerful manner.	84. Is cheerful	.31	66
21. Is talkative (in this situation).	4. Talkative individual	.30	42
6. Dominates the interaction.	14. Genuinely submissive (reverse keyed)	.29	47
48. Expresses self-pity or feelings of victimization.	78. Self-pitying	.28	46
62. Speaks quickly.	20. Has rapid personal tempo	.28	67
41. Keeps partner at a distance	48. Keeps people at a distance	.26	26
38. Is expressive in face, voice, or gestures.	43. Facially and/or gesturally expressive	.25	74
54. Speaks fluently and expresses ideas well.	98. Verbally fluent	.25	78
M		.31	58.20
	15 lowest items		
49. Seems interested in partner as member of opposite sex.	80. Interested in opposite sex	−.07	14
22. Expresses insecurity or sensitivity.	68. Basically nervous	−.04	46
23. Shows physical signs of tension or anxiety.	10. Anxiety and tension produce bodily symptoms	−.04	37
30. Seeks advice from partner (low = *partner seeks advice from subject*)	29. Is sought for advice (reverse keyed)	−.03	27
7. Appears to be relaxed and comfortable.	33. Is calm	−.01	60
44. Says or does interesting things (from partner's point of view)	57. Interesting, arresting person	−.01	29
45. Shows interest in conventional ways of judging people.	63. Judges self and others in conventional terms	.01	48
53. Offers advice to partner.	95. Gives advice	.05	32
17. Discusses unusually large number of topics.	3. Wide range of interests	.06	35
37. Behaves in a fearful or timid manner.	40. Generally fearful	.06	41

Table 9 (*continued*)

Behavioral Q-sort item	Personality Q-sort item	r	Operant/respondent score
27. Seeks reassurance from partner.	19. Seeks reassurance	.07	32
58. Demonstrates interest in topics related to power.	91. Power oriented	.08	43
33. Expresses warmth.	35. Has warmth; compassionate	.12	49
40. Expresses guilt (about anything).	47. Readiness to feel guilt	.12	41
47. Seems to view interaction as sexual encounter.	73. Eroticizes situations	.13	17
M		.03	36.73

Note. The behavioral Q-sort items are aggregated across three situations.

A slightly broader way to view the same result is to correlate the 62 operant/respondent scores with the two kinds of cross-situational consistency. Across all 62 items, operant/respondent scores correlated .51 ($p < .001$) with the average consistency of these BQ items across the three laboratory situations; among the 41 items that had analogues in the CQ, operant/respondent ratings correlated .54 ($p < .001$) with consistency between the lab and real life. These two correlations remained significant, being .35 and .48, respectively, *after* the influence of both the variability and reliability of the BQ items was partialed out of them. Therefore, although the operant/respondent distinction has something to do with item reliability (the correlation between operant/respondent scores and reliability is .32, $p < .05$), the relationship between this distinction and behavioral consistency is clearly substantive as well as psychometric.

Discussion

Level of Analysis and Behavioral Consistency

A widespread view in the personality literature for the past 20 years or so has been that the upper limit for correlation coefficients that reflect the consistency of people's behavior across any two even somewhat dissimilar situations is about .30 (Mischel, 1968), or perhaps .40 (Nisbett, 1980).[9] The basis for this view has been studies such as those by Hartshorne and May (1928) and Mischel and Peake (1982), who found their highest cross-situational correlations to be in this range.

This conclusion has not gone unchallenged. Several psychologists have pointed out that when aggregates of behavior, rather than single instances, are examined, cross-situational consistency correlations can climb dramatically (e.g., Epstein, 1979, 1983; Epstein & O'Brien, 1985; Jackson & Paunonen, 1985). Other writers have shown how the .30 or .40 limit is more respectable than many investigators have been aware. Effect size correlations between .30 and .40 characterize some of the most important experimental effects in the literature of social psychology (Funder & Ozer, 1983), and even a correlation of .30 is sufficient that a prediction made with this degree of validity is right about twice as often as it is wrong (Rosenthal & Rubin, 1979, 1982).

The present data would seem to make a more simple point: It no longer seems tenable to regard a correlation of .30 or even .40 as a ceiling for the consistency of unaggregated behaviors across different situations. Of the 62 cross-situational correlations calculated between Session 1 and Session 2 in the present study, fully 25 were greater than .40, and they ranged as high as .70. The other two cross-situational comparisons also revealed numerous behaviors with consistency better than .40.

These findings do not merely reveal consistency across identical situations. Sessions 1 and 2, across which so many behaviors were highly consistent, involved interactions with different, randomly selected partners and were held several weeks apart. Moreover, at Session 1, the laboratory situation was completely novel; by Session 2, subjects were more experienced. Perhaps most important of all, the subjects' behavior reflected this difference between the two situations. As we saw in Table 5, subjects were obviously much more relaxed and friendly, and much less nervous and irritable, in Session 2 than they had been in Session 1, among other differences.

So, why were the cross-situational consistency correlations so high? One apparent answer concerns the level of analysis at which behaviors were coded. We did not count speech disfluencies, instances of mumbling, or the number of dumb questions subjects asked their partners. Rather, we obtained reliable codings of the degree to which each subject "exhibited an awkward interpersonal style." The consistency of this latter behavior

[9] Because of occasional confusion about what these numbers have been claimed to represent, it may be worth quoting Nisbett at some length: "Predictability from one situation presumed by the investigator to tap a particular trait to another, phenotypically different situation (or behavior) presumed to tap the trait, virtually never exceeds .30, rarely exceeds .20, and most typically is in the range from .10 to .15. Let us be clear about what is meant by 'phenotypically different.' I mean by that term that at least one plausibly important element of the stimulus situation or behavioral setting, or of the 'morphology' of behavior, is different from one test or behavior occurrence to another" (1980, p. 116).

across Sessions 1 and 2 was .66. We did not count smiles, eye sparkles, or jokes but reliably coded "behaves in a cheerful manner." The consistency of this behavior across Sessions 1 and 2 was .60. And so forth.

Those personality psychologists who pioneered the empirical investigation of behavioral consistency usually conceptualized behavior at a very concrete level; as we pointed out earlier, typical behavioral measurements recorded whether a subject cheated at a specific game or how many homework assignments he or she turned in on time. Perhaps this is why the early studies of consistency reached such pessimistic conclusions. Behavior is simply not very consistent at this level of concreteness and specificity, or at least not consistent beyond about the .30 or .40 range. To obtain higher correlations than that, the present study suggests, one is well advised to conceive of behavior not in terms of its concrete manifestations, but in terms of its psychological meaning (see also Cairns & Green, 1979; Pettit, McClaskey, Brown, & Dodge, 1987).[10]

This conclusion has implications that go beyond how behavior could or should be coded. It can be viewed as speaking directly to the paradox that has received frequent attention in the personality literature and was specifically identified and discussed by Bem and Allen (1974) as

> the sharp discrepancy 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 (pp. 507–508).[11]

The results of the present study suggest one possible resolution: Intuitions and research have not been addressing the same phenomenon. Research has focused on the consistency of behavior. Intuition focuses on the consistency of personality. The two are related, but they are not exactly the same thing. When we regard an acquaintance as consistently fearful, or awkward, or cheerful, we are not necessarily expressing an intuition that all the behaviors by which one might manifest these dispositions will be consistent across all situations of the acquaintance's life. Rather, we are expressing a belief that in any situation in which these behaviors are relevant, our acquaintance will manage to exhibit fearfulness, awkwardness, or cheerfulness in some way or another. The number of different behaviors through which these and other dispositions might be manifested is nearly unlimited. But the integrative powers of the mind can allow human judgment to detect them, it would seem, with relatively little difficulty (Cairns & Green, 1979).

This conclusion implies that important topics for future research should be the specific behaviors by which personality dispositions are manifested, and the way that naive perceivers use these manifestations to make their own inferences about personality (e.g., Buss & Craik, 1983; Neisser, 1980; Riggio & Friedman, 1982, 1986; Scherer, 1978). In the meantime, the present results suggest that the resolution to the consistency paradox might be this: If behavior is defined only in terms of specific and concrete acts, then it indeed may not be particularly consistent across situations. But personality is.

Assessing Differences Between Situations

A long-standing, unresolved question in personality psychology has been how to assess the psychologically relevant properties of situations (Bem & Funder, 1978). The present study provides a demonstration of how one might begin to assess situations in terms of their effects on behavior. The average subject was less nervous in Session 2 than he or she was in Session 1, for example, and behaved in the debate in a fashion that was both more skeptical and more loud than he or she acted in either Session 1 or Session 2. These behavioral differences reflect, in a natural way, the operational differences that were built into these laboratory situations. The subjects were more experienced with the research setting in Session 2 than they were in Session 1; hence, apparently, they were more relaxed. The debate was designed to evoke argument, whereas Sessions 1 and 2 were unstructured opportunities to get acquainted; these differences were reflected by changes in friendly and combative behavior.

These findings suggest that one possible tactic for assessing the psychological differences between situations is to measure a large and representative set of behaviors in each of them. This technique is analogous to the procedure routinely used by experimental social psychology whenever it tests psychological hypotheses by comparing mean differences in a single behavior between two laboratory conditions. The present results suggest that by assessing multiple behaviors—as opposed to a single behavior—across a broad band, psychologists might come to a richer understanding of how situations differ from each other in their psychological and behavioral effects.

The second method by which the present study examined differences between situations was through examination of the consistency of behavior across them. This second method is essentially an extension of the one proposed by Bem and Funder (1978), who advocated assessing situations by examining the personality correlates of individual differences in behavior. However, none of Bem and Funder's subjects were observed in more than a single laboratory situation. Here, by contrast, each subject was observed in three situations, and analyses focused on the degree to which the subjects maintained their individual differences in behavior across them. The correlations revealed that behavioral consistency was more the rule than the exception—nearly all correlations were positive, many were significant, and some were rather large—but consistency was markedly greater between Session 1 and Session 2 than between either of these unstructured interactions and the debate. This finding buttressed the conclusion, reached through analyses of mean differences between situations, that Session 1 and 2 were more psychologically similar to each other than was Session 2 to the debate.

By the way, this conclusion was not necessarily expected. Sessions 1 and 2 were similarly structured, to be sure, but were separated by several weeks and involved interactions with different partners, whereas Session 2 and the debate were held a few minutes apart and involved interactions with the same

---

[10] We would of course not claim that all behaviors coded at the level of psychological meaning will exhibit significant cross-situational consistency. However, the present results suggest coding behaviors at a higher level of generality will make findings of consistency more likely.

[11] Bem and Allen (1974) went on to argue that our intuitions are correct because they generally use a trait term only when describing people for whom behavior is truly consistent on that dimension. We would agree, but here we will make a different point.

partner. A priori, it was not clear to us which two of these situations would be the more psychologically similar. The answer was obtained empirically, through two methods that supported each other's conclusions. One method focused on behavioral differences across situations; the other method focused on behavioral consistencies. Both approached the assessment of situations through an examination of their effects on behavior.

Cross-Situational Consistency Versus Cross-Situational Differences

The analyses just discussed assessed three situations through the complementary examination of two kinds of effect more usually viewed as being diametrically opposed to one another: behavioral differences and behavioral consistency. A common assumption is that the fact that people behave differently in different situations somehow implies that individual differences and personality are unimportant. But, as Ozer (1986) argued on theoretical grounds, situational specificity and behavioral consistency are not necessarily in opposition.[12] The present study provides an empirical demonstration. A large number of behaviors changed significantly, on average, across the three laboratory situations, but almost all the same behaviors also manifested positive correlations across these situations, and some of the correlations were impressively large. Moreover, across all 62 of the behaviors that were assessed, cross-situational differences and average cross-situational correlations were uncorrelated. For personality psychology, these findings lead to the following conclusion: Even though situations profoundly affect what people do, people can still manage to preserve their distinctive behavioral styles across situations.

This conclusion has implications for social psychology, as well. In research on attribution theory, dispositional explanations for behavior are often viewed as directly opposed to situational ones: Subjects are sometimes asked to rate the degree to which they think behaviors were caused by people or by situations (Funder, 1982). This practice reveals a misunderstanding exactly parallel to that evidenced by regarding effects of situations as antagonistic to effects of people. There is no reason why a person's behavior cannot both be determined by the situation and still be manifested to a consistent, relatively high or low degree by that person across the different situations of his or her life. Whether viewed as relevant to personality consistency or to attribution theory, therefore, the person–situation dichotomy may ultimately be a false dichotomy.

Differences Between Behaviors

The most surprising finding of this research may be its discovery of an apparently stable dimension of difference between behaviors, such that some behaviors are more consistent than others. This effect was demonstrated in two ways. First, the same behaviors that exhibited the highest consistency between any two of our laboratory situations also tended to be highly consistent across any two of the other situations as well: The correlations indexing the strength of this tendency all exceeded .70. Second, the behaviors that were most consistent across the three laboratory situations also tended to be more highly correlated with analogous items in the CQ, which reflected how the subjects' behavior in real life was evaluated by their friends and acquaintances. Also, the correlations indexing these tendencies remained significant after the influence of the differing variabilities and reliabilities of the BQ items was partialed out of them.

In an attempt to develop an explanation for this difference between behaviors, we turned to an idea expressed by David McClelland several years ago which was based, perhaps loosely, on a distinction offered more than half a century earlier by B. F. Skinner: the distinction between operants and respondents. This may have been an unfortunate move on our part. We already have been advised by several colleagues that these terms have acquired so much surplus meaning—spawning whole literatures on, for instance, the difference between operant and respondent (or classical) conditioning—that to raise them in the present context merely clouds and confuses matters. So, let us explicitly disavow any intended connection between the phenomena reported in this article and the vast literatures spawned over 50 years by Skinner's distinction between two kinds of behavior.

On the other hand, for the moment, we would also like to defend the fundamental importance of the distinction itself. Stripped of excess baggage, the distinction Skinner was trying to make back in 1931 was between behaviors that were *elicited* in obvious and direct response to specific situational stimuli and behaviors that did not require specific stimuli but were instead *emitted* by the organism across a wide range of situations. If this simple and basic distinction is accepted as valid, then McClelland's hypothesis, rather than being strange or radical in any way, suddenly becomes almost tautological. Behaviors that occur in response to specific stimuli are stimulus specific and therefore narrowly situation specific, by definition. Similarly, behaviors that do not require specific, eliciting stimuli are more likely to occur across a broad range of situations and to reflect properties—such as personality dispositions—of the people who emit them.

One way to view our results, therefore, is that we have shown how naive raters can identify, a priori, differences between behaviors that are and are not elicited by specific situational stimuli, as opposed to being emitted by and expressing characteristics of the behaving person. These ratings, in turn, to a significant degree predict which behaviors will manifest the least and the greatest correlations across situations. But, like Skinner himself, we have said almost nothing about why this difference between behaviors should exist. If one reads the behaviorist literature of 50 years ago, one sees a variety of attempts, none entirely successful, to develop a theoretical distinction between operants and respondents. In the end, many researchers of that era took Skinner's lead and eschewed further attempts to distinguish between the two kinds of behavior on principled grounds; instead, they accepted the distinction as purely descriptive and proceeded with their research from there.

Perhaps that is enough for now. The apparent fact that some

[12] Strictly speaking, they will be in opposition only when the situation is so powerful that it utterly obliterates individual differences—that is, a situation where everybody acts exactly the same. A few examples come to mind (e.g., a gunman enters a bank and yells "hands up!"), but they are rare. Nearly all situations in life, and certainly all of those examined in the present research, exhibit considerable "within-cell" behavioral variance.

behaviors are more consistent than others deserves to be the target of further investigation. A complete and satisfying explanation of this phenomenon and other matters pertaining to the cross-situational consistency of behavior probably must await not only further empirical research, but significant advances in the neglected field of personality theory.

References

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

Bakeman, R., & Gottman, J. M. (1986). *Observing interaction: An introduction to sequential analysis.* Cambridge, England: Cambridge University Press.

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. (1971). *Lives through time.* Berkeley, CA: Bancroft Books.

Block, J. (1978). *The Q-sort method in personality assessment and psychiatric research.* Palo Alto, CA: Consulting Psychologists Press.

Block, J. H., & Block, J. (1980). The role of ego-control and ego-resiliency in the organization of behavior. In W. A. Collins (Ed.), *Minnesota symposium on child psychology* (Vol. 13, pp. 39–101). Hillsdale, NJ: Erlbaum.

Bower, G. H., & Hilgard, E. R. (1981). *Theories of learning* (5th ed.). Englewood Cliffs, NJ: Prentice-Hall.

Buss, D. M., & Craik, K. H. (1983). The act frequency approach to personality. *Psychological Review, 90,* 105–126.

Cairns, R. B., & Green, J. A. (1979). How to assess personality and social patterns: Observations or ratings? In R. B. Cairns (Ed.), *The analysis of social interactions: Methods, issues, and illustrations* (pp. 209–226). Hillsdale, NJ: Erlbaum.

Chaplin, W. F., & Goldberg, L. R. (1985). A failure to replicate the Bem and Allen study of individual differences in cross-situational consistency. *Journal of Personality and Social Psychology, 47,* 1074–1090.

Colvin, C. R., & Funder, D. C. (in press). Predicting personality and behavior: A boundary on the acquaintanceship effect. *Journal of Personality and Social Psychology.*

Conley, J. J. (1985). Longitudinal stability of personality traits: A multitrait–multimethod–multioccasion analysis. *Journal of Personality and Social Psychology, 49,* 1266–1282.

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. (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.

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

Funder, D. C. (1982). On the accuracy of dispositional vs. situational attributions. *Social Cognition, 1,* 205–222.

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., & Colvin, C. R. (1988). Friends and strangers: Acquaintanceship, agreement, and the accuracy of personality judgment. *Journal of Personality and Social Psychology, 55,* 149–158.

Funder, D. C., & Dobroth, K. 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. (1986). 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.

Guilford, J. P., & Fruchter, B. (1978). *Fundamental statistics in psychology and education* (6th ed.). New York: McGraw-Hill.

Hartshorne, H., & May, M. A. (1928). *Studies in deceit.* New York: Macmillan.

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

Howard, A., & Bray, D. W. (1988). *Managerial lives in transition: Advancing age and changing times.* New York: Guilford Press.

Jackson, D. N., & Paunonen, S. V. (1985). Construct validity and the predictability of behavior. *Journal of Personality and Social Psychology, 49,* 554–570.

Kelly, E. L. (1955). Consistency of the adult personality. *American Psychologist, 10,* 659–681.

Kelly, G. A. (1955). *The psychology of personal constructs* (Vol. 1). New York: Norton.

Kenrick, D. T., & Funder, D. C. (1988). Profiting from controversy: Lessons from the person–situation debate. *American Psychologist, 43,* 23–34.

Lanning, K. (1988). Individual differences in scalability: An alternative conception of consistency from personality theory and measurement. *Journal of Personality and Social Psychology, 55,* 142–148.

Lay, C. (1977). Some notes on the concept of cross-situational consistency. In D. Magnusson & N. S. Endler (Eds.), *Personality at the crossroads* (pp. 143–146). Hillsdale, NJ: Erlbaum.

McClelland, D. C. (1984). Is personality consistent? In D. McClelland (Ed.), *Motives, personality and society* (pp. 185–211). New York: Praeger.

McKinnon, D. W. (1962). The nature and nurture of creative talent. *American Psychologist, 17,* 484–495.

McNemar, Q. (1947). Note on the sampling error of the differences between correlated proportions or percentages. *Psychometrika, 12,* 153–157.

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

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.

Murray, H. A. (1938). *Explorations in personality.* New York: Oxford University Press.

Neisser, U. (1980). On "social knowing." *Personality and Social Psychology Bulletin, 6,* 601–605.

Nisbett, R. (1980). The trait construct in lay and professional psychology. In L. Festinger (Ed.), *Retrospections on social psychology* (pp. 109–130). New York: Oxford University Press.

Ozer, D. J. (1986). *Consistency in personality: A methodological framework.* West Berlin, Federal Republic of Germany: Springer-Verlag.

Peck, R. F., & Havighurst, R. J. (1960). *The psychology of character development.* New York: Wiley.

Pettit, G. S., McClaskey, C. L., Brown, M. M., & Dodge, K. A. (1987). The generalizability of laboratory assessments of children's socially competent behavior in specific situations. *Behavioral Assessment, 9,* 81–96.

Riggio, R., & Friedman, H. S. (1982). The interrelationships of self-monitoring factors, personality traits, and nonverbal social skills. *Journal of Nonverbal Behavior, 7,* 33–45.

Riggio, R., & Friedman, H. S. (1986). Impression formation: The role of expressive behavior. *Journal of Personality and Social Psychology, 50,* 421–427.

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 effects. *Journal of Educational Psychology, 74,* 166–169.

Scherer, K. R. (1978). Personality inference from voice quality: The loud voice of extraversion. *European Journal of Social Psychology, 8,* 467–487.

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.

Skinner, B. F. (1931). The concept of the reflex in the description of behavior. *Journal of General Psychology, 5,* 427–458.

Skinner, B. F. (1966). *The behavior of organisms: An experimental analysis.* New York: Appleton-Century-Crofts. (Originally published 1938)

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).

Stephenson, W. (1953). *The study of behavior.* Chicago: University of Chicago Press.

Teitelbaum, P. (1977). Levels of integration of the operant. In W. K. Honig & J. E. R. Staddon (Eds.), *Handbook of operant behavior.* Englewood Cliffs, NJ: Prentice-Hall.

Terman, L. M., & Oden, M. H. (1947). *The gifted child grows up: Twenty-five years' follow-up of a superior group.* Stanford, CA: Stanford University Press.

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

Zuckerman, M., Koestner, R., DeBoy, T., Garcia, K. T., Maresca, B. C., & Sartoris, J. M. (1988). To predict some of the people some of the time: A reexamination of the moderator variable approach in personality theory. *Journal of Personality and Social Psychology, 54,* 1006–1019.

Appendix

Cross-Situational Consistency Correlations and Means for 62 BQ Items

	Cross-situational consistency correlations									Means and standard deviations					
	Total sample			Women			Men			Session 1		Session 2		Session 3	
BQ item	S1/S2	S1/D	S2/D	S1/S2	S1/D	S2/D	S1/S2	S1/D	S2/D	*M*	*SD*	*M*	*SD*	*M*	*SD*
1	.12	.15	.22	−.11	.18	.20	.30	.11	.24	5.61	1.71	5.35	1.54	4.87	1.09
2	.17	−.04	−.09	.13	−.05	−.06	.20	−.01	−.11	5.87	1.38	6.21	1.40	5.04	1.23
3	.27	−.05	.07	.28	.02	.02	.27	−.09	.11	4.38	1.38	4.56	1.41	5.01	0.84
4	.34	.09	.24	.27	−.10	.14	.40	.21	.33	6.21	1.35	6.39	1.28	6.44	1.14
5	.38	.13	.27	.41	.15	.27	.33	.09	.25	4.42	1.22	4.43	1.15	4.94	1.63
6	.27	.05	.32	.28	.01	.38	.25	.04	.24	4.77	1.61	4.76	1.45	4.90	1.73
7	.48	.24	.45	.48	.17	.39	.48	.30	.52	5.63	1.73	6.13	1.62	6.38	1.50
8	.58	.17	.18	.56	.29	.35	.59	.06	.03	6.04	1.49	6.46	1.33	6.27	1.20
9	.62	.41	.52	.64	.39	.50	.57	.41	.52	4.36	2.07	4.13	2.14	3.93	2.10
10	.63	.27	.43	.56	.25	.33	.63	.19	.46	5.71	1.47	5.54	1.49	5.36	1.70
11	.60	.30	.39	.39	.29	.33	.60	.25	.40	6.03	1.36	6.16	1.41	5.68	1.50
12	.39	.38	.47	.34	.28	.19	.43	.43	.61	3.86	1.47	3.56	1.35	3.97	1.46
13	.17	.06	.29	.09	.01	.26	.24	.10	.30	5.79	1.01	6.00	0.98	5.45	0.95
14	.66	.28	.31	.67	.35	.45	.62	.23	.21	4.11	1.94	3.60	1.72	3.87	1.82
15	.16	.04	.15	.04	−.08	.13	.28	.13	.19	4.71	0.87	4.68	0.69	4.84	1.31
16	.59	.47	.53	.55	.50	.62	.59	.41	.43	4.54	1.85	4.60	1.90	4.90	1.84
17	.10	.11	.05	.00	.10	.14	.16	.09	−.08	4.65	1.10	4.49	1.23	4.86	1.01
18	.38	.12	.20	.24	−.12	.09	.48	.33	.32	3.96	1.02	3.51	0.99	3.93	1.31
19	.38	.20	.23	.40	.26	.31	.35	.14	.11	5.45	0.94	5.43	0.98	4.72	1.52
20	.40	.10	.14	.34	.16	.21	.43	.07	.10	4.96	1.01	4.82	1.07	5.58	0.91
21	.38	.14	.21	.46	.29	.30	.32	−.03	.11	5.83	1.57	6.05	1.32	6.35	1.53
22	.56	.30	.38	.52	.32	.51	.60	.28	.24	4.71	1.25	4.49	1.11	4.29	1.24
23	.45	.28	.46	.40	.26	.52	.50	.30	.39	5.15	1.62	4.66	1.58	4.85	1.70
24	.44	.34	.29	.39	.29	.15	.48	.38	.40	5.36	0.84	5.24	0.72	5.58	1.22

(Appendix continues)

Appendix (*continued*)

	Cross-situational consistency correlations									Means and standard deviations					
	Total sample			Women			Men			Session 1		Session 2		Session 3	
BQ item	S1/S2	S1/D	S2/D	S1/S2	S1/D	S2/D	S1/S2	S1/D	S2/D	*M*	*SD*	*M*	*SD*	*M*	*SD*
25	.16	.12	.20	.07	.22	.12	.25	.02	.29	5.25	0.74	5.35	0.69	4.71	0.71
26	.41	.22	.26	.40	.35	.17	.40	.12	.30	5.33	1.14	5.38	1.19	4.88	1.24
27	.27	.17	.21	.34	.24	.19	.20	.08	.24	4.57	0.82	4.58	0.77	4.54	0.91
28	.47	.03	−.02	.56	.09	−.09	.40	−.01	.03	4.80	0.97	4.65	0.87	5.03	1.02
29	.11	.16	.34	.01	.11	.49	.20	.19	.23	6.14	0.93	6.20	1.01	5.61	1.00
30	.26	.09	−.04	.36	.21	.01	.17	−.05	−.17	4.69	1.20	4.71	1.14	4.87	0.67
31	.55	.43	.42	.55	.45	.45	.53	.41	.39	5.09	0.86	5.13	0.92	5.15	0.69
32	.43	.13	.30	.35	.24	.36	.48	.02	.27	3.78	1.00	3.60	0.90	4.00	1.18
33	.12	.05	.01	.00	−.03	.08	.23	.14	−.03	5.30	0.83	5.45	0.81	4.84	0.72
34	.10	.00	.09	−.08	.15	.19	.22	−.14	.00	4.09	0.66	4.06	0.73	4.18	0.73
35	.30	.02	.18	.31	.08	.19	.25	−.05	.18	4.26	0.90	4.24	0.93	4.51	1.14
36	.45	.47	.61	.29	.38	.45	.55	.52	.69	4.12	1.03	4.14	1.09	4.23	0.90
37	.65	.45	.60	.68	.52	.67	.62	.38	.54	3.88	1.70	3.64	1.64	3.39	1.61
38	.63	.52	.53	.65	.42	.56	.58	.58	.50	5.17	1.56	5.42	1.58	5.92	1.74
39	.04	−.05	−.02	.10	.02	.11	−.03	−.12	−.13	4.95	0.36	4.96	0.28	4.90	0.26
40	.08	.09	.12	.04	.00	.09	.09	.17	.14	4.54	0.57	4.47	0.70	4.30	0.68
41	.39	.23	.31	.29	.31	.29	.48	.17	.32	4.77	1.41	4.40	1.48	4.92	1.18
42	.36	.32	.19	.22	.18	.01	.46	.47	.33	5.56	1.00	5.43	0.85	5.89	1.08
43	.42	.15	.40	.27	.13	.53	.53	.19	.30	5.91	1.34	6.14	1.27	5.62	1.31
44	.17	.19	.14	.04	.14	.23	.29	.24	.04	5.78	1.00	6.08	0.97	5.98	0.94
45	.07	.16	.16	.02	.14	.04	.12	.19	.26	5.19	0.77	5.26	0.76	5.54	0.83
46	.15	.05	.12	.10	.04	.01	.23	.04	.23	5.53	1.17	5.34	0.95	4.80	0.42
47	.11	−.01	.06	.13	.04	.02	.09	−.04	.11	4.16	0.66	4.23	0.62	4.11	0.52
48	.28	.18	.27	.37	.24	.26	.22	.09	.31	4.68	0.85	4.80	1.05	4.24	0.66
49	−.03	−.02	.18	−.21	−.02	.20	.09	−.02	.16	4.31	0.83	4.44	0.75	4.20	0.57
50	.60	.28	.41	.52	.24	.34	.60	.25	.43	5.90	1.53	6.11	1.45	5.69	1.46
51	.05	.12	.06	.01	.10	.10	.08	.15	.03	4.97	0.55	4.93	0.38	6.18	1.21
52	.43	.48	.54	.39	.48	.39	.45	.48	.67	5.46	1.04	5.58	1.02	5.61	0.92
53	.29	.05	.01	.41	.06	−.02	.20	.04	.00	4.89	1.16	4.76	1.11	4.72	0.60
54	.42	.25	.14	.40	.26	.16	.41	.27	.15	6.00	1.30	6.25	1.17	6.23	1.75
55	.14	.08	.01	.19	.00	−.08	.08	.16	.09	4.91	0.88	4.91	0.85	4.66	0.36
56	.16	.09	.15	.00	.05	.16	.31	.12	.13	5.36	0.98	5.46	0.90	4.97	0.51
57	.70	.58	.65	.74	.55	.69	.67	.61	.61	4.11	1.39	4.30	1.56	4.94	1.57
58	.13	.17	−.05	.06	.17	−.05	.18	.17	−.06	5.00	0.67	4.92	0.51	5.50	0.66
59	.23	−.02	.19	.14	−.15	.21	.28	.10	.19	4.52	1.10	4.49	1.08	4.17	0.87
60	.57	.43	.60	.59	.40	.56	.53	.45	.67	6.09	1.58	6.37	1.58	6.35	1.72
61	.54	.20	.32	.44	.26	.48	.62	.15	.20	3.92	1.58	3.55	1.60	3.46	1.64
62	.59	.56	.53	.55	.45	.60	.60	.59	.55	4.88	1.17	4.99	1.13	5.12	1.38

Note. BQ = behavioral Q-sort; S1 = Session 1; S2 = Session 2; D = debate.

Received June 8, 1990
Revision received August 6, 1990
Accepted August 9, 1990 ■