Journal of Personality and Social Psychology
1984, Vol 47, No 4, 871-883

Temporal Stability and Cross-Situational Consistency of Affective, Behavioral, and Cognitive Responses

Ed Diener and Randy J. Larsen
University of Illinois

Consistency and stability of feelings were examined in reports that were completed on 3,512 occasions randomly sampled from the lives of 42 subjects. The stability and consistency of responses depended on the situations, individuals, and responses involved. High degrees of consistency were unusual for single responses, although mean levels of responding tended to be both highly stable and consistent. The consistency and stability of variables covaried, suggesting a connection between the two. Persons who were more consistent across one pair of situations tended to be more consistent across other situational pairs. The results indicate that the question of whether personality consistency exists does not have a simple answer, and requires knowledge of the persons, situations, responses, and level of analysis involved.

When Mischel (1968) raised the issue of whether behavior is cross-situationally consistent and stable for persons over time, a controversy that has continued for over a decade was initiated within the field of personality. As Underwood and Moore wrote in 1981, "The issue of consistency is one of the most controversial and most fundamental issues for the future of personality psychology. . . ." (p. 784). One major response to Mischel's critique has been the idea that personality theories do not require cross-situational consistency of behavior (e.g. Alker, 1972). Some personologists maintain that it is enough if behavior shows intrapsychic consistency. In other words, "Personality characteristics may be revealed in a variety of situations by different behaviors exemplifying the same trait" (Alker, 1972, p. 8). Another response to Mischel's thesis is that only some people are consistent on a particular trait (e.g., Bem & Allen, 1974). Yet a third response to Mischel, one that can be quite compatible with his social learning approach, is that personality should be studied idiographically. The idiographic approach suggests that each person may show a stable pattern of behavior across situations, but that he or she cannot be compared with others because the responses that covary will be unique to each individual. Finally, conceptualization of personality in terms of person by situation interactions (Ekehammar, 1974) is yet another response to Mischel. Interactional approaches, like situational ones, rest on the assumptions of a person's inconsistency of responses across situations (Endler & Magnusson, 1976). However, they are compatible with some theories of personality, such as Murray's (1938), that stress the interplay between personal and situational forces. Note that each of these defenses against Mischel's critique of personality rests on the admission that a person's responses are not cross-situationally consistent. It is our contention that knowledge about response consistency across a variety of samples, responses, and natural settings is necessary before this debate can be resolved in an intelligent manner. Not all approaches to personality require the degree of consistency that a strong trait view might imply. However, we believe that consistency data ought to be gathered across a variety of responses and situations before a firm commitment to one of the approaches is made.

Why are the questions of consistency and stability of responding so important? The

We wish to express our appreciation to Robert Emmons, Laura Spera, and Sharon Griffin for their help in the data collection and analyses. This research was supported in part by National Institute of Mental Health Grant No. MH-15140 for research training in personality and social ecology.

Requests for reprints should be sent to Ed Diener, Department of Psychology, University of Illinois, 603 E. Daniel, Champaign, Illinois 61820.

bedrock of the traditional idea of personality is that although individuals differ from one another, they show coherence in their behavior across time or place. Without such coherence there would be little point in a field that focuses on individuals. Instead, personologists would be better off studying momentary decisional processes, situational effects, and so forth, without reference to any ongoing stability in persons. However, even if one takes a coherence, interactional, or idiographic approach to personality, some patterned stability of individuals must be defended. In personological formulations, the person should at least be stable over time and across situations that are very similar.

One reason that the empirical base is not sufficient to completely answer questions about response consistency is that the answer is not likely to be simple. First, consistency and stability probably differ for different response domains. Even Mischel (1968) admitted that consistency probably occurs in the area of cognitive abilities and styles. Also, consistency can be conceptualized and measured in different ways. As already suggested, person consistency can occur either across situations ("consistency") or over time ("stability"). These two are related in that there is always some time difference when people are tested across situations and there is always some difference in the situations when people are tested over time.

Both consistency and stability can be considered either as correlations between single occasions or as the relation of responses between two situations or two periods of time based on the average of a number of occasions in each situation or time period. Responses on single occasions may not be particularly consistent because of a myriad of factors that may influence them. However, personal consistency may appear when responses are aggregated over several observations in a situation, and highly stable patterns may thus emerge. Jaccard (1974) showed that much higher predictability is gained when one is interested in such aggregates. Similarly, Epstein (1979) demonstrated across a number of response domains that reliability climbs as one aggregates more occasions over time. When one aggregates an individual's responses over occasions, there is a tacit admission that single respones are likely to be inconsistent. One is able to better predict a person's average responses over a number of occasions because the effect of moods, situations, and other factors is lessened or removed from what is being predicted. In other words, when data are aggregated for individuals, other factors that influence single responses will tend to be controlled because they will be averaged out over occasions.

Stability of responding over both short and long time periods has been found in a number of studies (e.g., Block, 1971; Block, Buss, Block, & Gjerde, 1981; Costa, McCrae, & Arenberg, 1980; Olweus, 1979; Schaie & Parham, 1976; Schuerger, Tait, & Tavernelli, 1982). Mischel (1979) recently wrote that there appears to be substantial temporal stability in behavior, but a small amount of cross-situational consistency. Indeed, he writes that the belief in personality traits comes about because of the perception of the stability in behavior (Mischel & Peake, 1982). However, there has been insufficient research on cross-situational consistency, especially in natural settings. For example, Koretzky, Kohn, & Jeger (1978), using independent raters, found moderate cross-situational consistencies in delinquents for apathetic (r = .42) and angry/defiant (r = .52) behaviors. On the other hand, Dudycha (1936) found only small amounts of consistency for punctuality measured across a variety of settings. The low consistency found by Dudycha was based, however, not on ratings but on measurement of single responses. Note that findings of higher cross-situational consistencies such as those based on peer ratings are usually based on data that are implicitly averaged across occasions. One reason that rater judgments may show relatively high cross-situational consistency (e.g., Koretzky et al., 1978) is that raters base their judgments on a number of occasions.

The purpose of the present study was to examine the temporal stability and cross-situational consistency of positive and negative affect, as well as several other cognitive and behavioral responses. We studied whether aggregated responses were consistent across situations and stable over time, and we also estimated the consistency and stability that would have been obtained if single responses

had been used. We also compared the stability and consistency of affect to (a) more cognitive judgmental feelings (e.g., life satisfaction), (b) feelings that denote behavioral predispositions or motivational propensities (e.g., sociability), and (c) broad behavioral responses (e.g., physical activity). We also examined consistency across differing types of situations. Another purpose of the present study was to compare different individuals in terms of consistency. If some individuals are highly consistent and others are highly inconsistent, then a totally nomothetic, individual-differences approach to affect is called into question.

It is our belief that science progresses by explaining regularities that are found in the world. We need to discover what responses are recurrent under what conditions in order to resolve the person/situation debate and move to more sophisticated theories of personality that are based on firm data. We are not advocating the kind of blind empiricism that has sometimes occurred in this area. However, we do advocate building a data base for a variety of response domains in order to determine and delimit the types of person consistency that may exist. We did not ask subjects to imagine how they might feel in a variety of situations. Rather, we assessed how they reported they felt on a number of occasions when in natural situations.

There is reason to believe that response consistency may be greater in the affective domain than in overt behavioral responses. Behavior is more subject to reinforcement contingencies, which may be highly idiosyncratic for specific situations. On the other hand, the everyday stereotype of moods is that they may also be quite variable. However, Epstein (1982) presented evidence that suggested that feelings show more temporal stability than either behavior or impulses. We examined not only positive and negative affect over time in our respondents' everyday lives, but also the consistency of affect across specific situations.

A question relevant to the topic of response consistency is also a question that has plagued interactionists: namely, how to define situations. Debates have raged over whether situations should be defined subjectively or objectively, and also about how molar or molecular the situations should be (Moos & Insel, 1974). We have selected for study situations that differ in a number of ways. Our first situational dimension was degree of interaction with other people, varying from social interaction to being alone. This dimension can be objectively defined and seems to be one of the most fundamental and important ways in which situations can be categorized. The second situational contrast we used was work versus recreation. Although this issue is somewhat more subjective, there is fairly widespread agreement in our culture about what is work and what is play. In addition, this too seems to be an important situational subdivision in that there are very different expectancies and rewards in our culture for these two situations. Last, we used the situational typology of novel versus typical as subjectively judged by the person. We thought that this more abstract dimension might be relevant to the question of consistency because a person may develop a habitual way of responding to typical but not to novel situations. As can be seen, our situations depended more on culturally shared meanings and activity and less on the physical environment per se.

In summary, the present study was aimed at collecting ecologically valid data on the feelings of individuals across time and across situations. We hoped to determine the degree of stability and consistency for (a) various feelings and responses, (b) single versus aggregated responses, (c) different types of situations, and (d) the whole sample versus certain individuals who were particularly consistent or inconsistent.

Method

Participants

The subjects were 42 University of Illinois undergraduate students, equally divided between men and women. The participants were enrolled in an independent study course on life satisfaction in which they received a grade of satisfactory or unsatisfactory. The subjects were quite heterogeneous in terms of motivation for enrolling in the course, in personality, and in their seriousness as students. There were virtually no constraints on what university students could enroll in if they were interested.

Procedures

Participants completed mood forms at two random times every day for a six-week period. Each student wore

a watch with an alarm that had been preset to go off at random times during the individuals' waking hours. The random times, which were based on sampling without replacement, were generated so that every 10-min period during the waking day was covered over the six weeks. Times were also selected so that the first one in a day would come in the morning or early afternoon, with the second one occurring in the later afternoon or evening. The subject would set his or her alarm according to this list for the next alarm time on each occasion that the alarm had just sounded. Several precautions were taken so that the alarm would not be expected, thus making the measurements nonreactive. For example, the next time the alarm was to go off was several hours to many hours later, and so the subject would have forgotten about it by that time. It was the experience of most subjects that they would think about the alarm for the first day or two, but thereafter the alarm would catch them unexpectedly.

When the alarm sounded, the subject was to focus on the feelings of that instant and then complete the mood questionnaire immediately. If it was impossible to do so (e.g., during a test), the student could postpone completing the form up until one hour later. If it was still impossible to complete the form, an extra random time later in the day was to replace the missed time. Under no circumstances were the mood forms to be completed several hours or a day later because memory distortion was thought to be a potential problem. In order to ensure compliance, students had to turn in the forms every day.

Mood Form

The mood form consisted of 41 self-report scales about feelings. Twenty-three of the scales were monopolar and answered on a scale ranging from 0 (*I felt it not at all*) to 6 (*extremely much*). Mood adjectives covered both positive affect (e.g., happy, joyful, and pleased) and negative affect (e.g., unhappy, frustrated, and depressed). Positive and negative affect are treated separately throughout this report because we and others have found that the two vary independently across persons (Bradburn, 1969; Diener & Emmons, in press; Zevon & Tellegen, 1982). When the results were analyzed, four adjectives were averaged to provide positive affect (happy, joyful, pleased, and enjoyment) and five negative adjectives were averaged for negative affect (depressed, unhappy, frustrated, angry, and worried/anxious). These items were selected based on factor analytic work in other studies we have conducted. Additional scales included adjective self-ratings such as "Productive" and "Satisfied with my life." There were also a series of bipolar adjective pairs (rated on a 9-point scale) such as: "crabby–cheerful," "physically active–inactive," "feeling ill–feeling well," "unaroused–aroused," "tired/lethargic–energetic," "see world as beautiful/good–ugly/bad," "low self-esteem–high self-esteem," and "sociable–want to be alone." Subjects responded to each bipolar pair on a 9-point scale; the midpoint (5) was labeled *Does not apply* and the two extremes (1 and 9) were anchored with *Very much* Subjects were to indicate how much they were feeling one or the other of the adjectives contained in each bipolar pair when their beeper went off.

For this report, in addition to the more purely affective adjectives, self-reports were also chosen to reflect several other response domains: cognitive judgmental (e.g., life satisfaction), motivational (e.g., sociability), and behavioral (e.g., physical activity). Thus comparisons across domains could be made and the relative degree of stability and consistency could be more readily judged.

Situations

Also reported on the mood form was the current activity in which the person was engaged. Subjects indicated whether they were working, recreating, or maintaining (e.g., eating, walking to class). We focused on the work and recreation situations for this report because they were seen as most dissimilar. Although subjects might indicate that they were in more than one situation at once (e.g., eating while studying), these times were excluded from the data analyses because they led to nonindependence among situations. The second situation concerned sociality (degree of interaction with others). The most social situation was being involved in verbal conversation; next were the categories semisocial and social presence; last was being alone, not in the presence of others. In our analyses we again have focused on the two most dissimilar situations, social interaction versus being alone. The third situational dimension was more subjective: the degree of novelty in the situation. Subjects rated the situation as typical or novel on a 9-point bipolar scale. For purposes of data analysis, we excluded times that were marked at the midpoint of 5, and dichotomized the scale into two categories.

Artifact Checks

In addition to the Crowne–Marlowe scale (Crowne & Marlowe, 1964) measuring social desirability, we included two measures to indicate how each subject used the number system of the scales. A response style of using high or low numbers to report one's affect could potentially inflate consistency and stability estimates. The artifact measures indicated how intense of an emotional response each subject meant when he or she responded with a 2, a 4, and so on. On one measure, subjects described in detail how they felt when they marked, say, either a 2 or a 4 on positive and on negative affect separately. These four descriptions were then rated on a −100 to 100 scale by two independent raters who showed high interrater agreement for both the positive and negative descriptions (*rs* = .90 and .92). These ratings indicated the degree to which subjects used the number scale in a conservative or liberal direction in describing their feelings. After they were averaged across the two raters, the responses (2 and 4) for positive affect were summed, as was also done for negative affect. A high score on the positive sum and negative sum indicated responding in a conservative fashion on the scale, so that the scale numbers meant more extreme affect for the subject. In another effort to detect artifactual responding, we had subjects indicate where each of their scale numbers from 0 through 6 would be on a line. The line was marked continuously with positive affect words that had been prescaled for intensity and were placed at the appropriate position on the line. This provided subjects another opportunity to indicate the intensity of feeling they meant by their number responses. Thus we could check on the possibility

that stability and consistency might be due to response style rather than affect per se.

Results

The first analyses considered the degree to which self-report artifacts may have biased the findings. Specifically, did social desirability or idiosyncratic use of the number scales inflate the consistency found? The first answer is that the degree of consistency varied so greatly among variables, with some being quite inconsistent, that it seems unlikely that artifacts had a strong effect. The artifact measures showed substantial correlations with each other (average $r = .63$) and low correlations with the Crowne–Marlowe (average $r = .18$), suggesting that some subjects did report using the number scales in characteristic ways. The ratings between these potential artifacts and subjects' response ratings were also quite modest. The artifact checks correlated an average of $r = .17$ with positive affect and an average of $r = .04$ with negative affect. The Crowne–Marlowe correlated $-.01$ with positive affect and $-.34$ with negative affect.

The Crowne–Marlowe score and the three scale usage scores were partialed out of key consistency and stability correlations. The three-week stability coefficients changed from .79 to .76 for positive affect and from .81 to .77 for negative affect. The consistency correlations also changed very little. These analyses indicated that response artifacts had very little substantive influence on the results or conclusions of this article and they are not considered further. Because response bias appeared to have so little impact on the findings, we thought it inadvisable to correct the data for response style variance. In addition, it is problematical whether such "artifacts" are totally spurious or whether they represent substantive differences between individuals. We did, however, correct the findings for another potential problem, measurement unreliability.

One difficulty in estimating true consistency and stability is that the measures may contain some degree of unreliability or error. For example, subjects may make numerical errors when reporting their behavior. Consistency and stability estimates can be corrected for such unreliability. In order to estimate the reliability of our measures, we have correlated (both over all occasions and also within situations) the average of individuals' responses from odd and even days over the six-week period and these are shown in Table 1. The average number of occasions in each situation are also shown. Other things being equal, scores based on larger aggregates show higher reliabilities. These odd/even-day correlations are good estimates of the reliability of the measures because they are based on a large number of observations within the same settings during the same time period. Internal consistency reliabilities (coefficient alpha) were also computed for the affect measures because they were composed of several items (Cronbach, 1951). The coefficients were .89 and .84, based on the full sample of moments for positive and negative affect, respectively. One can see that both types of reliability are acceptably high. In this report we present results that are both uncorrected and corrected for unreliability (based on the odd–even reliability estimates). One may desire to correct for measurement unreliability in order to estimate the true-score correlation between two measures (Ghiselli, Campbell, & Zedeck, 1981). However, such corrections may cloud substantive issues. For example, when the focus of concern is on the consistency and stability of responding, the reliability of that responding is the substantive issue that is being explored. Because it is impossible to totally separate unreliability that is due to one's measures from the lack of stability in the phenomenon, we have presented both the corrected and uncorrected correlations.

Table 2 shows the correlations for individuals' responses across pairs of situations. The correlations corrected for unreliability (Lord & Novick, 1968) are also shown. One can see that some variables show much greater cross-situational consistency than others, regardless of whether corrected or uncorrected correlations are examined. For example, life satisfaction shows very high consistency across all situation pairs, whereas the feeling of sociability is invariably inconsistent. For some variables such as activity, productivity, and cheerfulness, the degree of consistency depends very much on which situational pairs are being considered. Thus these results suggest clearly that the issue of consistency should not be treated in a global manner.

Table 1
Odd–Even Reliabilities for Responses Within Situations

Response	Situation pairs						
	All	Social	Alone	Novel	Typical	Work	Recreation
Affect							
Positive	.83	.88	.74	.85	.94	.91	.91
Negative	.87	.87	.86	.82	.97	.94	.90
Bodily feelings							
Aroused	.86	.86	.76	.63	.83	.88	.86
Energetic	.92	.76	.69	.58	.68	.70	.73
Feeling well	.80	.86	.82	.77	.95	.89	.82
Behavior							
Physically active	.88	.90	.64	.82	.72	.86	.88
Productive	.92	.82	.64	.76	.82	.84	.89
Behavioral predispositions							
Sociable	.96	.74	.75	.66	.88	.82	.80
Cheerful	.95	.76	.76	.82	.90	.81	.77
Cognitive/judgmental							
World beautiful	.95	.94	.84	.92	.97	.93	.89
Self-esteem	.84	.82	.71	.86	.94	.82	.74
Satisfied with life	.88	.97	.97	.95	.99	.98	.99

Note. Average number of occasions in each situation was as follows: Social = 25, alone = 19, novel = 24, typical = 57, work = 28, recreation = 28.

Situations were selected in order to increase their dissimilarity. If a variance components analysis were applied to these data, then using the extremes of the situational distributions would attenuate the person effects. This is not a problem in the present correl-

Table 2
Aggregated Cross-Situational Consistency Correlations

Response	Uncorrected			Corrected for unreliability		
	Social–alone	Novel–typical	Work–recreational	Social–alone	Novel–typical	Work–recreational
Affect						
Positive	.58	.67	.70	.72	.75	.77
Negative	.70	.80	.74	.81	.90	.81
Bodily feelings						
Aroused	.50	.56	.60	.62	.77	.69
Energetic	.24	.43	.34	.33	.69	.47
Feeling well	.57	.60	.77	.68	.70	.90
Behavior						
Physically active	.11	.58	.47	.15	.76	.54
Productive	.29	.56	.16	.40	.71	.19
Behavioral predispositions						
Sociable	.10	.15	.01	.13	.20	.01
Cheerful	.13	.47	.37	.17	.55	.47
Cognitive/judgmental						
World beautiful	.80	.84	.78	.90	.89	.86
Self-esteem	.41	.62	.58	.54	.69	.74
Satisfied with life	.92	.96	.97	.94	.99	.98

Table 3
Corrected Cross-Situational Consistencies Disaggregated To Single Occasions

Response	Situational pairs		
	Social–alone	Novel–typical	Work–recreation
Affect			
Positive	.10	.08	.11
Negative	.16	.19	.13
Bodily feelings			
Aroused	.07	.08	.07
Energetic	.02	.06	.03
Feeling well	.09	.06	.25
Behavior			
Physically active	.01	.08	.04
Productive	.03	.06	.01
Behavioral predispositions			
Sociable	.01	.01	.00
Cheerful	.01	.03	.03
Cognitive/judgmental			
World beautiful	.30	.18	.18
Self-esteem	.05	.06	.09
Satisfied with life	.44	.70	.67

tional analyses. However, to the extent that there is a linear person–situation interaction, selection of situational extremes will lead to an underestimation of consistency. If the interaction is nonlinear, the impact of selecting situational extremes would depend on the nature of the nonlinear function.

In Table 3 we present the consistency correlations disaggregated to single occasions. We have disaggregated the correlations based on the average number of occasions the subjects were in each situation, using a formula presented by Ghiselli et al. (1981). These figures enable one to estimate the average value one would expect to obtain if two individual random moments were selected for each individual and then correlated. The most striking thing, of course, is the low value of most of these correlations. Only life satisfaction shows consistency correlations across single occasions that most would consider substantial. All other variables show very low cross-situational consistency correlations when single occasions are considered.

In Table 4 we present the temporal stability correlations computed across all situations. The first column represents the correlations

Table 4
Temporal Stabilities Across All Situations

Response	Three-week aggregated uncorrected	Three-week aggregated corrected	Three-week disaggregated corrected
Affect			
Positive	.79	.95	.48
Negative	.81	.93	.39
Bodily feelings			
Aroused	.81	.95	.45
Energetic	.65	.70	.10
Feeling well	.44	.55	.05
Behavior			
Physically active	.58	.66	.08
Productive	.65	.71	.10
Behavioral predispositions			
Sociable	.34	.35	.03
Cheerful	.76	.80	.16
Cognitive/judgmental			
World beautiful	.89	.93	.40
Self esteem	.70	.84	.20
Satisfied with life	.87	.99	.87

Table 5
Corrected Three-Week Stabilities Within Situations

Response	Situations					
	Social	Alone	Novel	Typical	Work	Recreation
Affect						
Positive	.79	.75	.61	.85	.58	.83
Negative	.72	.88	.48	.86	.69	.72
Bodily feelings						
Aroused	.68	.55	.68	.94	.57	.86
Energetic	.32	.94	.71	.86	.73	.65
Feeling well	.05	.50	.41	.49	.19	.65
Behavior						
Physically active	.51	.60	.46	.82	.56	.54
Productive	.73	.84	.36	.72	.75	.69
Behavioral predispositions						
Sociable	.49	.71	.29	.62	.29	.64
Cheerful	.70	.70	.38	.93	.27	.90
Cognitive/judgmental						
World beautiful	.85	.57	.71	.91	.83	.89
Self-esteem	.51	.70	.39	.81	.61	.62
Satisfied with life	.93	.91	.90	.93	.91	.93

for each variable when subjects' average scores for the first three weeks of the study are correlated with their average scores for the second three weeks. One can see that the temporal stability of different responses also varies greatly. A number of variables show high stability, but the stability of feel well/ill and feelings of sociability are quite modest. The middle column contains the three-week stability coefficients that have been corrected for the total odd–even unreliability estimates. The last column presents the values from the middle column that have been disaggregated down to single-occasion estimates of stability by a reverse application of the Spearman–Brown formula (Allen & Yen, 1979). When one considers the disaggregated figures, it is apparent that many variables are also not very stable when only single occasions are considered. Nevertheless, one can see that with the exception of life satisfaction, there is only modest stability for single responses on the other variables.

One can see the result of considering stability within particular situations by referring to the values in Table 5. Life satisfaction is stable within every situation. Affect tends to be quite stable over time in typical situations, but only moderately stable in novel and work situations. Although one may notice greater stabilities for all variables in typical situations than in novel ones, one must remember that the number of observations is much smaller for novel situations.

When one compares the situational stabilities in Table 5 with the cross-situational consistencies presented in Table 2, it is difficult to draw general conclusions about whether there is more consistency or stability in aggregated behavior. However, one can also compare the disaggregated consistencies shown in Table 3 with the disaggregated stabilities shown in Table 4, thus correcting for differences in the number of observations. This comparison suggests that affect, arousal, and the cognitive variables show greater stability than they do consistency. Nonetheless, the stability of many variables between single occasions is very small (e.g., feeling well/ill and physical activity). The correlations for feeling well/ill were undoubtedly attenuated by ceiling effects because most subjects felt well most of the time.

The covariation between stability and consistency can also be examined. Are those responses that are most stable the same ones

as those that are most consistent? After we created the positive and negative sums, there were 32 self-rating variables. The stability and consistency correlations for each were converted to Z scores and these were correlated across variables. The remarkably high correlation of .81 ($p < .001$) indicates that there was a strong tendency for variables that were inconsistent to also be those that were unstable. Is this simply due to the reliability of the measures? One can note from Table 1 that the overall reliability of the measures was similar and high. Thus differences in scale reliability cannot explain why the measures that are inconsistent also tend to be those that are unstable. Clearly, the variables that show coherence within individuals across time also tend to show it across situations.

Were some individuals more consistent across situations? Absolute value difference scores were taken between each pair of situations for each variable. For example, we computed the absolute value of average positive affect in social situations minus average positive affect in "alone" situations. This was done for each variable for each of the three pairs of situations. Such scores represent how much each individual changed from situation to situation. These change scores were then correlated between situational dimensions for each variable separately, and these correlations are shown in Table 6. A caveat is worth mentioning: Change scores are often highly unreliable even when the measures themselves are quite reliable. Thus correlations based on change scores can be highly attenuated (McNemar, 1969). Therefore, we computed the reliability of the change scores based on the formula suggested by Allen and Yen (1979). These reliability estimates for the change scores, although somewhat lower than the reliabilities for the component scores, were still substantial, averaging .61 across all variables. Nevertheless, the figures in Table 6 should be used to give a rough idea of whether some individuals are more consistent than others, without giving a precise estimate of the strength of this effect. It appears that there is a tendency for all variables for the more consistent people to be more consistent across all three situational dimensions. One should note that for variables such as life satisfaction, which are highly reliable, virtually all individuals must be consistent relative to one another. Therefore, some individuals must be virtually unchanging on such a variable, with others changing slightly for person consistency differences to emerge. Although the correlations in Table 6 should not be taken as precise values, it should be noted that the correlations were all positive and definitely are not distributed around zero. This indicates that there is a tendency for some individuals to be more consistent than others across various situational pairs. This appeared to be true for all variables. Thus some individuals appear to be more consistent and others less so.

Table 6
Correlations of Absolute Difference Scores

Response	Difference Score Pairs: Novel–typical & social–alone	Social–alone & work–recreation	Novel–typical & work–recreation
Affect			
Positive	.25	.50	.41
Negative	.40	.51	.22
Bodily feelings			
Aroused	.19	.54	.35
Energetic	.00	.16	.42
Feeling well	.21	.24	.54
Behavior			
Physically active	.41	.33	.33
Productive	.46	.23	.12
Behavioral predispositions			
Sociable	.14	.68	.40
Cheerful	.40	.36	.53
Cognitive/judgmental			
World beautiful	.43	.34	.43
Self-esteem	.21	.14	.15
Satisfied with life	.18	.36	.46

Discussion

In our study we sampled affect in subjects' everyday lives. Each subject completed approximately 84 mood reports that were scattered over a six-week period. In all we have a total of 3,512 occasions on which moods and other responses were reported. Because the reports were completed at random times,

they represent an ecological sample of affect and a representative sample of the situations in which participants spent their time. In summary, our major findings were as follows:

1. The degree of consistency not only differed for different responses and sets of situations, but specific responses were more or less consistent in different sets of situations. This finding suggests that general statements about consistency may be overly simple. Similarly, arranging situations on a single similarity dimension may be too simple because different aspects of situations may influence different responses. Within-situation stability was also diverse, depending on the response and situation involved.

2. Despite the diversity in consistency just referred to, several general trends can be noted. Life satisfaction was invariably stable and consistent, whereas feelings of sociability were not so. Affect and arousal showed moderately strong stability. Nevertheless, even after aggregation some responses were very consistent and others were not. Thus person effects can be weak in some cases even if long-term trends are examined. In other cases, person effects for long-term means can be quite strong.

3. Aggregating data across occasions resulted in much higher stability and consistency estimates than those based on disaggregated estimates. Such aggregation is analogous to randomization and experimental control in laboratory research because aggregation helps reduce the influence of uncontrolled factors. There were striking differences between the correlations for aggregated data and those disaggregated to single occasions. This indicates that there are consistent and stable long-term trends in mean levels of responding for individuals, but in general single responses will show very low levels of consistency and stability. Recall that our responses were randomly sampled from the everyday lives of our respondents. The low estimated single occasion correlations suggest short-term fluctuations or variability in the responses of our subjects, even though there is a long-term trend in the mean level of their responses.

4. It appeared that most variables tended to show somewhat greater stability that consistency. This difference was most pronounced for affect, arousal, and the cognitive variables. However, each of these also showed strong consistency when aggregated data were analyzed. Only one variable, feeling well/ill, showed somewhat more consistency than stability.

5. There has recently been interest in whether some individuals are more consistent than others (e.g., Campus, 1974; Chaplin & Goldberg, in press; Kenrick & Braver, 1982; Rushton, Jackson, & Paunonen, 1981). In our data it appeared that some individuals were more consistent across all types of situations than were others. Because every single correlation between change in one situational pair and another situational pair was positive, this is a strong indication that some individuals are consistently more consistent than others.

6. Stability and consistency strongly covaried across responses, indicating that a sharp demarcation cannot be drawn between the two. Although stability and consistency imply lack of change along the dimensions of time and situations, each always includes the other to some degree. Cross-situational consistency always occurs over time and temporal stability occurs over situations because no two situations can ever be truly identical. Those responses that were most stable over time were the same ones that were most consistent across situations. We cannot account for this finding by differences in the reliability of the measures. Mischel (1979) has stressed that stability exists in behavior because of the temporal stability of reinforcement contingencies, but that inconsistency across situations exists because persons recognize fine differences in the reinforcement systems of differing situations. The high correlations between stability and consistency indicate that certain additional factors must work to keep some responses relatively unchanging. Other responses appear to be more free to vary both across time and situation. Placing the locus of stability entirely in the environment fails to account for the high correlation found between the stability and consistency of variables. It is probable that some variables are more internally stable, whereas others are more reactive to environmental change.

There are several important implications of the aggregation data. On the prediction side, one can expect much stronger results

when aggregated data are being used because situational and other effects tend to cancel out. Of course, aggregation across time or across persons has been used for decades in other areas of psychology in order to average out factors in which the scientist was not interested. However, there are potential problems with aggregation that should be noted. First, it is easy to lose sight of the fact that in focusing on a particular variable, one will be ignoring and averaging out other variables that also influence the behavior being studied. This can be a curse or a blessing depending on the purpose of the study and the phenomenon in which one is interested. Second, when aggregation is used it becomes more difficult to compare effect sizes. Because aggregation over greater numbers of incidents will almost inevitably lead to larger correlations, the size of the relations found will depend both on the number of occasions aggregated and on the underlying effect size per se. In other words, a correlation may be large simply because it is based on a larger aggregate, not because the underlying connection is larger. This makes effect size comparisons between variables and between studies difficult because the aggregated number of occasions usually differ. This problem is especially serious for personality correlations based on ratings made by knowledgeable informants because their ratings depend on different numbers of observations. One approach to solving the problem related to varying numbers of observations is to disaggregate to single responses as we have done and as discussed by Golding (1975).

On the conceptual side, the high degree of stability and consistency obtained when occasions are aggregated reaffirms the fact that behavior (as well as other responses such as affect) tends to be complexly determined. In predicting a single occasion, many factors, including uncontrolled ones, will influence a response. In most cases, nomothetic person effects are not so strong in determining single responses that they outweigh other factors. There are exceptions. Some cognitive judgments, at least once initially formed, tend to be very stable and consistent even for single occasions. Although for most types of responses persons are usually only mildly consistent across situations, the person consistency that does exist may be very interesting in terms of psychological theory. As Diener (1983) has pointed out, a recurrent phenomenon such as person consistency may be theoretically important regardless of effect size. The statistical strength of a relation is only one factor among many when its theoretical importance is judged. Even correlations of only .05 could be important in some cases. The fact that some level of nomothetic consistency exists for virtually all variables is a phenomenon requiring explanation. The present data reveal that small correlations across individual occasions can translate into large correlations if one is interested in long-term trends in behavior. Nevertheless, it should always be recalled that certain other important influences on single responses are averaged out if one aggregates the behavior of individuals. It is one thing to note that personality correlations for responses on single occasions are low, but quite another to maintain that they are zero. Virtually all studies and all the consistency correlations we presented are positive. They are not distributed around zero. Thus the likelihood that consistency is actually zero as proposed by Bem (1972) is extremely small.

If one is interested in long-term average levels of behavior, strong person effects can be found for many variables. This approach to personality has been advocated by Jaccard (1974), Epstein (1979, 1982) and more recently by Buss and Craik (1983). What is suggested is that there are stable and consistent mean differences in the behavior of individuals, but also variability around these mean levels so that individual responses are not strongly predicted by person differences. If one is interested in the single responses of individuals within specific situations, as is Mischel (Mischel & Peake, 1982), then nomothetic personality may not be very potent for most responses, although variables such as life satisfaction are an exception. The high correlations obtained in this study and by Epstein (1979, 1982) for aggregates do not mean that personality is any stronger in its ability to explain individual responses. However, if one is interested in long-term mean levels of behavior, there may be person effects that are strong, although some variables may show person effects that are small even at this level (e.g., feeling sociable). Which level of analysis a scientist chooses to study will

depend on the phenomena he or she is attempting to explain. The understanding of long-term person effects is a justifiable undertaking. When investigators aggregate data for persons, they will amplify person effects by averaging out short-term situational effects, as well as random errors of measurement (Diener & Larsen, 1984). The explanation of specific behaviors in specific situations may also be justifiable but will undoubtedly require many variables. If one is interested in understanding momentary moods or emotion, the personality of the respondent would only be one small influence among many. However, the investigator might wish to understand long-term differences in happiness, in which case personality might play a larger role. Ultimately the goal must be to develop well-grounded theories to explain phenomena, and not simply debate the question of the "strength" of personality.

Because few behavioral responses show high consistency across single occasions, there is a temptation to use idiographic and interactional approaches. Whether these will be more successful in predicting single occasion behavior or whether these effects will also be overlaid by a large number of other factors has yet to be seen. Certainly in the domain of overt behavior in which response consistency is low, other approaches should be tried, but they do not necessarily have a better a priori probability of success.

One major conclusion that one can draw from the present study is that person consistency varies greatly, depending on the response domain, the situations being considered, and the particular persons involved. Given this complexity, it seems advisable to quit debating whether person consistency exists, and to begin exploring the factors that control consistency (e.g., Monson, Hesley, & Chernick, 1982; Snyder, 1982). At the level of theory, the degree of consistency that does exist suggests a place for nomothetic person variables, but also indicates that more refined theories must be developed. We need to determine why and when consistency and inconsistency occur (Diener, 1983). In this regard, we need to carefully consider why some variables, some situational dimensions, and some individuals tend to show larger amounts of consistency. However, we should not expect an overly simple answer to the question of when consistency will occcur.

References

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

Allen, M. J., & Yen, W. M. (1979). *Introduction to measurement theory* Monterey, CA: Brooks/Cole.

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

Bem, D. J., & 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.

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

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

Bradburn, N. (1969). *The structure of psychological well-being* Chicago: Aldine.

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

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

Chaplin, W. F., & Goldberg, L. R. (in press). Predicting some of the people some of the time, with some of the data: Results of a large-scale replication and extension of the Bem and Allen study of individual differences in cross-situational consistency. *Journal of Personality and Social Psychology*

Costa, P. T., & McCrae, R. R., & Arenberg, D. (1980). Enduring dispositions in adult males. *Journal of Personality and Social Psychology, 38,* 793–800.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16,* 297–334.

Crowne, D. P., & Marlowe, D. (1964). *The approval motive: Studies in evaluative independence* New York: Wiley.

Diener, E. (1983). *The phenomenon of person consistency.* Unpublished manuscript, University of Illinois.

Diener, E., & Emmons, R. (in press). The independence of negative and positive affect. *Journal of Personality and Social Psychology.*

Diener, E., & Larsen, R. J. (1984). *Limitations of aggregation* Unpublished manuscript, University of Illinois.

Dudycha, G. J. (1936). An objective study of punctuality in relation to personality and achievement. *Archives of Psychology, 204,* 1–39.

Ekehammar, B. (1974). Interactionism in personality from a historical perspective. *Psychological Bulletin, 81,* 1026–1048.

Endler, N. S., & Magnusson, D. (1976). Toward an interactional psychology of personality. *Psychological Bulletin, 83,* 956–974.

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. (1982). A research paradigm for the study of personality and emotions. In M. M. Page (Ed.), *Nebraska symposium on motivation 1981* (pp. 91–154). Lincoln: University of Nebraska Press.

Ghiselli, E. E., Campbell, J. P., & Zedeck, S. (1981). *Measurement theory for the behavioral sciences.* San Francisco: Freeman.

Golding, S. L. (1975). Flies in the ointment: Methodological problems in the analysis of the percentage of variance due to persons and situations. *Psychological Bulletin, 82,* 278–288.

Jaccard, J. J. (1974). Predicting social behavior from personality traits. *Journal of Research in Personality, 7,* 358–367.

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

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

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores.* Reading, MA: Addison-Wesley.

McNemar, Q. (1969). *Psychological statistics* New York: Wiley.

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

Mischel, W. (1979). On the interface of cognition and personality: Beyond the person–situation debate. *American Psychologist, 34,* 740–754.

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

Monson, T. C., Hesley, J. W., & Chernick, L. (1982). Specifying when personality traits can and cannot predict behavior: An alternative to abandoning the attempt to predict single-act criteria. *Journal of Personality and Social Psychology, 43,* 385–399.

Moos, R. H., & Insel, P. M. (1974). *Issues in social ecology* Palo Alto, CA: National Press Books.

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

Olweus, D. (1979). Stability of aggressive reaction patterns in males: A review. *Psychological Bulletin, 86,* 852–875.

Rushton, J. P., Jackson, D. N., & Paunonen, S. V. (1981). Personality: Nomothetic or idiographic? A response to Kenrick and Stringfield. *Psychological Review, 88,* 582–589.

Schaie, K. W., & Parham, I. A. (1976). Stability of adult personality: Fact or fable? *Journal of Personality and Social Psychology, 34,* 146–158.

Schuerger, J. M., Tait, E., & Tavernelli, M. (1982). Temporal stability of personality by questionnaire. *Journal of Personality and Social Psychology, 43,* 176–182.

Snyder, M. (1982). *Understanding individuals and their social worlds.* Paper presented at the Annual Convention of the American Psychological Association, Washington, DC.

Underwood, B., & Moore, B. S. (1981). Sources of behavioral consistency. *Journal of Personality and Social Psychology, 40,* 780–785.

Zevon, M. A., & Tellegen, A. (1982). The structure of mood change: An idiographic/nomothetic analysis. *Journal of Personality and Social Psychology, 43,* 111–122.

Received March 29, 1983
Revision received June 10, 1983 ■