Reprinted from: CHILD DEVELOPMENT, 1978, 49, 483-494.

The Reliability and Stability of Individual Differences in Infant-Mother Attachment

Everett Waters

University of British Columbia

50 infants were seen twice in the Ainsworth and Wittig Strange Situation to assess individual differences in the quality of infant-mother attachment at 12 and at 18 months of age. Evidence for the stability of individual differences was clearly a function of the level of analysis. The reliability of discrete-behavior variables was typically very low, and there was little evidence of temporal stability. There was clear evidence for stable individual differences in the analysis of behavior category data. This was especially true of behavior toward the mother during reunion after brief separations. Classification data based largely on reunion behavior and crying were even more stable across the 6-month interval. Each infant was assigned to 1 of 3 categories (secure/normative, avoidant, or ambivalent) on the basis of the patterning of attachment behavior at 12 months. 48 of the 50 infants were independently reassigned to the same category on the basis of the same behaviors at 18 months. In contrast to time sampling of phenotypically similar discrete behaviors, assessments which take into account the behavioral context of behavior yield more reliable assessments of individual differences in the quality of infant-adult relationships.

Infant-adult ties have often been conceptualized in terms of an underlying causal or dispositional trait. Some infants are said to be strongly attached to an adult; others to be less strongly or not at all attached (e.g., Feldman & Ingham 1975) . Attachment is often operationalized in terms of a small number of "attachment behaviors." These include behaviors which promote proximity to presumed attachment figures and behaviors which are perhaps directed more often toward attachment figures than toward nonattachment figures. To operationalize attachment in terms of attachment behaviors is to say that approaching, touching, looking, clinging, protesting separation, etc., are valid "indices" or measures of attachment. Strong attachments have often been inferred from (as well as used to explain) Performance of attachment behavior at high frequency and/or intensity. Trait models assume that various attachment indices are significantly correlated and that individual differences in these behaviors are stable across time.

The attachment construct has recently fallen under the cloud of a general dissatisfaction with the use of trait constructs and individual differences research strategies in personality and developmental psychology (e.g., Masters & Wellman 1974). The most influential critiques have been narrowly empirical; that is, they have not asked whether trait constructs are the kinds of constructs we want to build attachment theory around. Instead, recent critiques have dwelt upon the evidence that attachment behaviors are not strongly intercorrelated and are

This research was supported in part by Program Project Grant 5 POl HD05027 from the National Institute of Child Health and Human Development (NICHHD) and by a training grant (NICHHD, 5-2 HD00105), both to the Institute of Child Development University of Minnesota. The author's appreciation is extended to Drs. Mary Ainsworth and Mary Main for their generous assistance, and to Drs. L. Alan Sroufe, Dante Cicchetti, Leah Matas, and Brian E. Vaughn, each of whom contributed many hours of invaluable assistance in the analysis of the present data. This research was submitted to the Graduate School of the University of Minnesota in partial fulfillment of the requirements for the degree Doctor of Philosophy.

not remarkably stable across time. From this evidence it has been concluded that the concept. of attachment is seriously lacking in construct validity.

Negative evidence in construct validation research can indicate (a) that the measures employed have not measured the construct reliably, *(b)* that the experimental design failed to test (or poorly instantiated) the theory, or (c) that the body of theory surrounding the construct is incorrect (Cronbach & Meehl 1955) . Negative evidence in construct validation research can be taken as valid negative evidence only when failures of measurement and design (a and *b* above) can be ruled out. The present research was not undertaken in defense of trait conceptualizations of the attachment construct (see Sroufe & Waters [1977a] and Waters [Note 1] for relevant critiques of trait models) . It was undertaken to assess the adequacy of typical assessments of individual differences in attachment behavior. In addition, the research was designed to test the hypothesis that success in the search for stable individual differences is in part a function of the level of organization at which behavioral individuality is assessed.

Method

Subjects

Fifty infants (25 males, 25 females) and their mothers participated in the experiment. They were recruited from a subject pool maintained at the Institute of Child Development. Birth announcements in Minneapolis-St. Paul newspapers initiated written contact with families. Those who returned preaddressed postagefree cards indicating interest in participating in research were included in the subject pool. Fifty families were contacted by telephone and visited in their homes for explanation of the present research. The mothers were all between 22 and 30 years of age, and all of the families were intact. All but three of the infants studied were firstborn. Socioeconomic status of the families covered the lower middle to upper middle classes.

Design

Each infant and its mother was seen in the Ainsworth and Wittig (1969) Strange Situation within 2 weeks of the infant's first birthday and again when it was 18 months old. Four analyses were involved in the experiment: (a) 12-18-month correlations among time samples of discrete behaviors and estimates of the reliability of the time sample data, (b) 12-18-month correlations among rated behavior categories, (c) an analysis of the stability of patterns of rated behavior categories (using a classification system described in Ainsworth, Blehar, Waters, & Wall (1978), and (d) an analysis of the effects of error variance on the stability of these classifications, using artificial data generated from the 12- and 18-month data described above.

Procedure

The Strange Situation is a laboratory procedure developed by Ainsworth and Wittig (1969) to highlight the operation of an attachment behavioral system in the year-old infant. The procedure consists of eight episodes presented in a standard order for all subjects. All subjects were tested in an 11 *X* 14-foot room at the Institute of Child Development. The room was equipped with two chairs (one for the mother and one for the stranger), magazines for the mother, and a variety of age-appropriate toys for the infant (puzzles, push toys, dolls, etc). A brief summary of the procedure is presented in table 1. The order of episodes is so arranged that the infant experiences a series of increasingly (mildly) stressful situations (new room, unfamiliar adult, separation from mother but in company of an adult, separation, and alone).

The procedure was designed to be comparable to brief everyday experiences common in our culture. The goal was to facilitate observation of heightened attachment behavior in such conditions in order to better understand its function and organization and to highlight individual differences. Ainsworth, Bell, and Stayton (1971) have shown a clear relationship between behavior in the Strange Situation and the operation of an attachment-exploration balance in the home. The relationship seems strong enough to justify the use of this procedure to predict such behavior in 1-year-olds. The Strange Situation was not designed to determine the onset or strength of attachment to an adult (cf. Feldman & Ingham 1975); it presupposes that an attachment already exists. The procedure is not well suited to subjects under the age of 1 year. Without evaluation of its relationship to the attachment-exploration balance in older children, the Strange Situation as analyzed in the present experiment should be interpreted with caution if subjects are 2 years old or older. A detailed discussion of the use of the Strange Situation is presented by Ainsworth et al. (1978).

Response Measures

Each Strange Situation was videotaped, and all response measures were scored from these records. Experience showed that recording with a

TABLE 1
Summary of the Strange Situation Procedure

Episode	Persons Present	Time	Events and Procedures
1	M, B	Variable (approx. 1 min)	M and B are introduced into S/S room by E. If necessary, M interests B in toys before being seated. M does not initiate interaction but is responsive to bids from B
2	M, B	3 min	M remains seated and is responsive to bids for interaction but does not initiate
3	M, B, S	3 min	S enters and is seated; sits silently for 1 min; talks to M for 1 min; engages B in interaction and/or toy play for 1 min
4	B, S	3 min (less if B extremely distressed)	M leaves room, S allows B to play alone but remains responsive to interactive bids. If B is crying, S offers contact and tries to comfort. If B refuses or resists, S does not persist. Terminate episode after 1 min hard crying or on M's request
5	M, B	3 min	M calls B from outside door and steps inside, pausing at doorway to greet B and to reach and offer contact. If necessary, B is held and comforted then reinterested in toys; otherwise, M is seated and re mains responsive to bids from B but does not initiate
6	B	3 min (less if B extremely distressed)	M leaves room; B remains alone. Terminate episode if 1 min hard crying ensues or on M's request
7	B, S	3 min (less if B extremely distressed)	S returns and is seated. If B is crying or begins to cry without pause, S offers contact and tries to comfort. If B cannot be comforted and crying continues (or on M's request), terminate episode
8	M, B	3 min	M calls B from outside door and steps inside, pausing at doorway to greet B and to reach and offer contact. If necessary, B is held and comforted and then reinterested in toys; otherwise M is seated and remains responsive to bids from B but does not initiate if B is content in toy play

NOTE.-M - mother; B - baby; S - stranger.

pivot-mounted camera from a single position was sufficient in most cases and that concurrent dictation of a narrative record was largely redundant with information available from the videotape alone.

Two types of behavioral measures were scored-time samples (number of 10-sec intervals in which a behavior occurred) and ratings of categories of behavior. Scores were prorated proportionally for curtailed episodes.

Time samples of discrete behaviors and crying.-The frequency (or duration) with which looking (2 sec or longer), glancing (less than 2 sec), vocalizing, smiling, gesturing, approaching, touching, and holding were directed toward each adult was estimated for 30 randomly selected subjects in terms of time samples. The number of 10-sec intervals in which the behavior occurred was recorded for each episode. The number of 10-sec intervals in which crying occurred was recorded for each episode for all 50 subjects. The Pearson correlation between individual total scores for two independent raters for episodes 3, 4, and 8 was greater than .80 for each variable, and there were no mean differences between raters.

Ratings of behavior categories.-Ainsworth has developed a series of behavioral variables scored from the Strange Situation in the form of seven-point ratings assigned to behavior toward each adult separately in each episode. These measures were developed to take into account the fact that many attachment behaviors serve common goals (e.g., approaching, reaching, and vocalizing can each have the predictable outcome of achieving proximity to an adult). These measures attempt to take into account the behavioral and situational context in which a given behavior occurs (e.g., a delayed approach on reunion contributes to a lower score than an immediate approach). Since the significant aspects of timing, intensity, and context are

judged by Ainsworth to differ from episode to episode, when behavior is directed toward the mother as opposed to the stranger, and when different combinations of behaviors are involved, a rating format rather than a weighted composite was adopted. The variables scored were:

1. Proximity seeking (PS). The intensity and persistence of the baby's efforts to gain (or regain) physical contact (or more weakly, proximity) with an adult.

2. Contact maintaining (CM). The degree of activity and persistence in the baby's efforts to maintain physical contact with an adult once he has gained it (especially such active resistance to being released as clinging or protesting); also, behaviors such as sinking in while held which tend to delay the adult's attempts to release the baby (i.e., to prolong contact by not signaling readiness for release).

3. Proximity and interaction avoiding (PA). The intensity, persistence, duration, and promptness of any active avoidance of proximity or interaction, even across a distance, especially in reunion episodes. Included here are aborted approaches upon reunion, turning the face away when greeted, prolonged pout and refusal to make eye contact or to interact, and mild signs of wariness of the stranger accompanied by retreat to the mother. This rating does not include behavior which denotes only active interest in toys by an infant who is not distressed by separation or by the presence of a stranger (see Sroufe & Waters [1977] for a discussion of heart-rate data as a tool in validating the distinction between active avoidance and distraction or preoccupation).

4. Contact resisting (CR). The intensity and frequency or duration of negative behavior evoked by g person who comes into contact or proximity with the baby, especially behavior accompanied by signs of anger. Relevant behaviors include pushing away, dropping or hitting toys offered, body movements in resistance to being held. More diffuse indications include tantrums, and especially a prolonged pout or cranky fussing or other signs of inability to be comforted by contact with the adult. The behavior may alternate with active efforts to achieve or maintain contact, and both can be scored high in the same episode.

5. Distance interaction (DI). Spontaneous indications of positive interest in an adult, in the absence of proximity. Includes smiling, vocalizing, gestures, and play carried out with some attempts to elicit the adults interest or interaction.

Each of these behaviors was scored on a 1-7 scale on which at least every odd point is anchored to one or more specific patterns of response. The anchoring descriptions were selected by Ainsworth from typed transcripts of the actual behavior of 1-year-olds in the Strange Situation, and in this respect the scales have a clear descriptive advantage over rating scales anchored to adjective descriptors. The complete set of scales is included as an appendix in Ainsworth et al. (1978) .

Since the results of this experiment were expected to reflect upon the usefulness of the Strange Situation procedure, efforts were made to insure that the scoring of these variables was in accordance with the practices of Ainsworth and her associates. Dr. Mary B. Main generously scored episodes 5 and 8 from a number of our videotapes. The correlations between her independent scoring of 15 randomly selected subjects and ours were high enough to indicate substantial agreement in all but one case (r - .97, .87, .77, .84, and .61 for PS, CM, CR, PA, and DI, respectively).[1] While the correlation for DI was significant beyond the .01 level, the absolute value is relatively low, and results for this variable should be viewed accordingly. Pearson correlations among independent rescoring of episodes 3, 7, and 8 for 25 randomly selected video records were greater than .80 for each variable, and there were no significant mean differences between raters.

Classification

In addition to the time sampling and behavior ratings mentioned above, each infant's behavior in the Strange Situation was summarized by assigning it a category designation on the basis of patterns of the rated interactive behavior categories and the crying data. The classification scheme used was developed by Ainsworth and her colleagues and has been used widely to assess the quality of infant-mother relationships (see Ainsworth et. al. 1978 for a review). The three major categories and their relationships to interactive behavior and crying are summarized in table 2. Group B is the modal classification for middleclass 1-year-olds; in past research approximately 65% of a sample were typically placed in this group. Subgroup B3 is the largest subgroup (typically 40% of total sample) and is the group

[1]For those interested in establishing agreement with criterial scorings of these scales a set of prescored 1/2-inch videotapes from a sample of 12- and 18-month-olds in the Strange Situation is available on loan by arrangement with the author.

showing the most effective use of the mother as a secure base from which to explore, both at home and in the Strange Situation. Groups A and C are typically smaller (approximately 20% and 15% of total sample, respectively), and together the groups are often termed "anxiously attached."

Because the subgroups of each classification are small and have not been well studied, the emphasis in the present experiment was upon the larger groups A, B, and C. Nonetheless, all infants were assigned to subgroups, in part because attention to subgroup differences greatly facilitates agreement as to classification assignments. Each infant was classified at 12 and 18 months by independent judges. Classifications were not based on recorded scores alone but on the video record as a whole. Ninety percent of the subjects were classified by two or more judges at both ages. Interjudge agreement as to classification into the A, B, and C groups was 91%, 94%, and 85%, respectively; overall subgroup agreement was 84%. Disagreements were conferenced and resolved by reference to group and subgroup means provided by Ainsworth et al. (1978) .

Since the results of the present experiment were expected to reflect upon the usefulness of Ainsworth's Strange Situation classification system, efforts were made to insure that classifications were made in accordance with the practices of Ainsworth and her colleagues. Data on rated interactive behavior and crying from all episodes from a sample of 105 subjects provided by Ainsworth were employed in a multiple discriminant-function analysis of the A, B, C groups. The resulting discriminant functions were used to develop classification equations which were applied to the present 12- and 18-month data to obtain empirical classifications as similar as possible to those that a criterial judge would have assigned to our subjects. Despite the fact that the sample provided by Ainsworth was relatively small for this analysis, empirical classifications agreed with our own classifications approximately as well as the classification functions have been shown to cross-validate on a subsample of the Ainsworth data (68% A, 90% B, 35% C). Empirical classifications which differed from ours were consistently in the direction of classifying our A and C infants in group B. This also occurs on cross-validation of Ainsworth's data on a subgroup from the same, sample and is to be expected as a result of the small size of the A and C groups available for developing classification functions. These results, along with the distribution of subjects assigned to each group at 12 months (20% A, 60% B, 20% C) and the fact that group means on interactive behaviors and crying did not diverge greatly from means reported by Ainsworth et al. (in press), provide support for the conclusion that our classifications largely corresponded to the criteria developed by Ainsworth and her colleagues.

Results

The results of the analyses of discrete behaviors, interactive behavior categories, classification, and the analysis of the effects of error variance on classification are presented below.[2] In order to increase the reliability of individual scores, the data for each variable were summed across episodes as follows: preseparation behavior toward mother (episodes 1+2+3) reunion behavior toward mother (episodes 5 + 8), preseparation behavior toward stranger (episode 3), and behavior toward stranger during separation (episodes 4 +7). Crying data were combined into three composite scores, preseparation (episodes 2+3), separation (episodes 4 +6 + 7) , and reunion (episodes 5 + 8).

Time Samples of Discrete Behaviors

The 12-18-month correlations between each discrete behavior as it was directed to the mother or stranger in preseparation, separation, and reunion episodes are presented in table 3. Even if we discount episode 3 because the stranger's behavior as well as her initial unfamiliarity might reasonably reduce temporal stability for that episode, only four of the remaining 21 correlations reach conventional significance levels. These results are consistent with the data reviewed by Masters and Wellman (1974) in which there were consistently very few signs of stability of discrete behaviors, regardless of whether the intervening interval was 3 min, 1 day, 4 months, or longer.

[2] Since normative data on Strange Situation behavior have been reported extensively, they are not repeated for this sample. The only significant age effects or trends indicate that 18 month-olds are more mobile more vocal, and perhaps slightly less distressed in the Strange Situation than they were as 12-month-olds. The only significant sex effects indicated that crying (and its correlates) was greater in males than in females in the second separation and reunion sequence (episodes 7 and 8) at both ages. Descriptive statistics for the present sample are available from the author on request.

While these results appear to be valid negative evidence against the hypothesis that attachment behavior is stable across time, this can only be so if the possibility of inadequate assessment can be ruled out. One approach to this issue is afforded by the conventional psychometric theory of test reliability (Cronbach 1951; Nunnally 1967; Wiggins 1973). If we consider each 10-sec sampling interval to be a test item which is passed or failed (the target behavior occurs or does not occur), and consider each 3 (6, 7) -min episode a test consisting of 18 (36, 42) such items, we can compute an index of the reliability of individual scores (Cronbach's α) from item statistics.[3] Since an assumption of time-sampling methods is that the target behavior is equally likely to occur in any sampling interval, we can make use of the simplifying assumption that each of our "items" is of equal difficulty (equally likely to be passed).

The α reliabilities of each discrete-behavior score for behavior toward mother and stranger are presented in table 4. Since the reliabilities of the present time samples of discrete behaviors could be increased to any desired level by increasing the number of 10-sec intervals of observation (i.e., by increasing test length), Spearman-Brown estimates of the duration of time sampling necessary to achieve conventional psychometric standards of reliability are also presented for each behavior at both ages in table 4. The results indicate that the reliability of typical Strange Situation assessments of individual scores. (as opposed to group means) for discrete behaviors is frequently too low to support an adequate test of the temporal stability of these behaviors (or to evaluate correlations with other behaviors), especially when data from individual 3-min episodes are used.

Interactive Behavior and Crying

The 12-18-month correlations between each rated interactive behavior category for the entire sample are presented in table 5. Since crying is also a category of behavior rather than a discrete behavioral act, 12-18 month correlations for crying during preseparation, separation, and reunion episodes are also presented in table 5. If we discount episode 3 because the stranger's behavior as well as her initial unfamiliarity might reasonably reduce temporal stability in that episode, 13 of the remaining 18 correlations reach conventional significance levels. These results clearly contrast with the results for discrete behaviors, reported above. Evidence for temporal stability is especially clear among behaviors directed toward the mother in the reunion episodes and for crying, the behaviors that are most important in the classifications discussed below, as indicated in table 2.

Table 2
Summary of Strange Situation Classifications

	Classification Criteria (From Reunion Episodes 5 And 8)[a]				
Classification Descriptor	Proximity Seeking	Contact Maintaining	Proximity Avoiding	Contact Resisting	Crying
A (2 subgroups) . "Avoidant"	Low	Low	High	Low	Low (preseparation), high or low (separation), low (reunion)
B (4 subgroups) . "Secure"	High	High (if distressed)	Low	Low	Low (preseparation), high or low (separation), low (reunion)
C (2 subgroups) . "Ambivalent"	High	High (often preseparation)	Low	High	Occasionally (preseparation), high (separation), moderate to high (reunion)

[a] Typical of the group as a whole; subgroups differ in nonreunion episodes and to some extent in reunion behavior. See Ainsworth et al.

[3] A familiar approach to the measurement of test reliability is the method of intercorrelating split-halves of the test, using sums of odd- and even-numbered items. Split-half correlations are essentially instantaneous test-retest reliabilities when they are adjusted upward (using the Spearman-Brown formula) to account for the fact that each of the correlated tests is only half as long as the total test. Of course a test can be divided into halves in a number of different ways. Cronbach's α is equal to the mean of all possible corrected split-half correlations. It is also equal to the familiar Kuder-Richardson reliability estimate (KR-21 in the case of items of equal difficulty).

Table 3

Correlations Between 12- And 18-Month Time Samples Of Discrete Attachment Behaviors

	Target Adult			
	Mother		Stranger	
	Presepa ration (7 min)	Reunion (6 min)	Presepa- ration (3 min)	Sepa ration (6 min)
Look or glance	.070	.220	- .050	.110
Vocalize.	.360*	-.071	.121	.240
Smile	.462**	-.160	.200	.682**
Gesture.	.120	- .110	- .087	-.100
Approach	-.153	.040	.113	.090
Touch	.444**	.110	-	.260
Hold on	.260	-	-	-.080

Note. N = 30; dashes indicate that a behavior did not occur at one age.

Duration of combined episodes.
* y < .05 (one-tailed test).
** p < .01 (one-tailed test).

Classification

The 12- and 18-month classification data for each subject are presented in table 6. Overall, 48 subjects were classified in the same A, B, C group at 12 and 18 months; 30 subjects were classified in the same subgroup at 12 and 18 months. Cohen's (1960) index of nominal scale agreement (κ) was computed and tested as described by Fliess, Cohen, and Everitt (1969) for both group and subgroup classifications. It is computed by correcting the observed rate of agreement (same classification at both ages) for the rate of agreement expected by chance alone.[4] The κ *'s* for both A, B, C classification *(*κ = .92) and for subgroup *(*κ = .53) are significant beyond the .001 level (z=8.81, $p<10^{-10}$ and z=9.34, $p< 10^{-10}$, respectively). While ratings of interactive behavior categories and crying scores were substantially more stable from 12 to 18 months than time samples of discrete behaviors, classifications based on profiles or patterns of interactive behavior and crying were even more stable.

Effects of Random Error on Classification

While the interactive behaviors and crying data upon which classifications were based proved to be significantly correlated across the 12-18-month interval, even an average correlation of .50 implies substantial variation in individual scores and profiles. Since the A, B, C classifications showed substantially more temporal stability than the individual scores upon which they were based, there may be patterns of consistency within the residual 12-18-month variance of the individual variables that are not reflected in table 5. One approach to this question would be to compare reliability estimates similar to those presented in table 4 with the 12-18-month correlations in table 5. If the reliabilities were not substantially above the observed temporal stability estimates, then the best explanation of the unexplained 12-18-month variance would be that it is largely random error.[5] Unfortunately, internal consistency reliability estimates cannot be computed for the type of rating scales used in the present experiment, and Ainsworth et al. (1978) have demonstrated that independent short-term test-retest assessments are not feasible because of carryover from initial testing. Pairwise correlations among behavior categories also seem poorly suited to the problem of evaluating residual variance, especially if patterns of behavior rather than total scores are of interest.

The following auxiliary analysis employed the A, B, C classification system to help determine whether the unexplained variance in rated interactive behavior categories and crying scores is best interpreted as random (error) variance. Three sets of interactive behavior and crying data were involved in the analysis: (a) data on 155 12-month-old subjects (105 provided by Ainsworth plus the data from the 12-month testing on the present sample), (b) data from the 50 subjects tested at 18 months in the

[4] Given the observed marginal frequencies for groups A, B, and C at 12 and 18 months (20%, 60%, 20%, and 18% 64%, 18%, respectively), the rate of 12-18-month consistency expected by chance alone is 46%. The observed rate of agreement was 96%; κ - O (observed) - E (expected) /1 - E.

[5] It would not be highly desirable for the A, B, C classification to be insensitive to large amounts of random variation in individual scores (a) because this would suggest that the categories are perhaps so broad that the evidence for stability is unimportant, and (b) because this would be more consistent with the hypothesis that the classifications tap underlying individual differences in temperamental variables than with the hypothesis that they are useful in describing individual differences in the organization of attachment behavior.

Table 4
Reliability Of Time-Sampled Discrete-Behavior Data

Variable	Reliability Of Time-Sampled Behavior (Cronbach's α)		Spearman-Brown Estimate Of Time Sample (Min) Necessary To Achieve α=.90	
	12 Months	18 Months	12 Months	18 Months
Behavior toward mother:				
Preseparation episodes 2 and 3 (7 min):				
Look and glance	.51 (.28)	.51 (.28)	60.8	60.3
Vocalize	.45 (.24)	.66 (.39)	77.3	33.2
Smile	.53 (.29)	-	56.1	-
Gesture	.03 (.01)	.49 (.27)	1,905.8	65.1
Approach	.35 (.13)	.51 (.28)	116.0	59.8
Touch	.47 (.25)	.25 (.12)	71.0	184.1
Hold on		.71 (.44)	-	25.4
Reunion episodes 5 and 8 (6 min):				
Look and glance	.62 (.45)	.79 (.65)	32.7	14.5
Vocalize	.60 (.43)	.71 (.55)	36.3	21.7
Smile	-.26 (.15)	-	157.7	
Gesture	.02 (.01)	.37 (.23)	2,802.6	92.8
Approach	.78 (.64)	.43 (.27)	14.9	70.7
Touch	.48 (.32)	.36 (.22)	58.5	97.7
Hold on	.95 (.91)	.86 (. 75)	2.7	8.9
Behavior toward stranger:				
Preseparation episode 3 (3 min):				
Look and glance	.81	.21	6.2	102.8
Vocalize	.57	.66	14.0	14.0
Smile		-	-	- -
Gesture	.53		24.1	-
Approach	.51	.69	25.6	12.4
Touch		-	-	- -
Hold on		-	-	- -
Separation episodes 4 and 7 (6 min)				
Look and glance	.85 (.74)	.69 (.53)	9.9	24.2
Vocalize	.66 (.49)	.71 (.55)	27.9	21.9
Smile	.60 (.43)	.44 (.28)	36.5	68.8
Gesture	-.73 (.58)	-	20.3	
Approach	.26 (.15)	.53 (.36)	151.3	47.2
Touch	.70 (.54)	.72 (.56)	23.3	20.7
Hold on	.95 (.91)	.96 (.92)	2.6	2.1

Note.-Dashes indicate reliability is 0.0 or that a behavior did not occur at one age. Spearman-Brown estimates of the reliability of scores based on 3-min episodes are given in parentheses.

present experiment, and (c) a set of artificial data generated from the 12-month data of the present experiment. The artificial data were generated independently for each variable starting with the actual 12-month data of 50 subjects. Error variance was added to individual scores such that the correlation between the actual data and the artificial data was equal to the 12-18-month correlation for each variable. The resulting data simulated the effects of error variance on true scores (12 month Strange Situation data) in that changes in one subject's score on one variable were uncorrelated with changes on other variables and that variable means were not changed and intercorrelations among variables were attenuated.

Table 5
Correlations Between 12- And 18-Month Ratings Of Interactive Behavior Categories

	Target Adult: Mother		Target Adult: Stranger	
	Preseparation (4 Min)	Reunion (6 Min)	Preseparation (3 Min)	Separation (6 Min)
Proximity seeking	.423**	.303*	.033	.286*
Contact maintaining	.720**	.300*	-.020	.320*
Proximity avoiding -		.621**	.207	.229
Contact resisting		.508**	-.056	.274
Distance interaction . . .	.065	.308*	.180	.319*

	Preseparation (7 Min)	Separation (9 Min)	Reunion (6 Min)
Crying	.765**	.411**	.425**

Note. -N e 50; Dashes indicate that the behavior did not occur at one age.
* p < .05 (one-tailed test).

The data from 155 12-month-old subjects were employed in a multiple-discriminant function analysis of the A, B, C groups. The resulting discriminant functions were used to develop classification equations which were used to classify the subjects in this development group and were also applied to the data from the 50 18-month-old subjects of the present experiment and to 10 independent sets of artificial data generated as described above. The classification results for the 18-month-old sample and for the artificial data are presented in table 7.

Table 6
Classifications Based On Patterns Of Interactive Behavior And Crying

	18 Month							
12 Month	A1	A2	B1	B2	B3	B4	C1	C2
A1	**1**	**3**	0	0	0	0	0	0
A2	**2**	**3**	0	0	1	0	0	0
B1	0	0	**2**	**0**	**1**	**0**	0	0
B2	0	0	**2**	**6**	**1**	**0**	0	0
B3	0	0	**2**	**1**	**8**	**0**	0	0
B4	0	0	**1**	**0**	**0**	**6**	0	0
C1	0	0	0	0	0	0	**3**	**1**
C1	0	0	0	1	0	0	**4**	**1**

Both the 18-month data and the artificial data yielded significant cross-validation results, $\kappa = .596$ ($z = 5.32$, $p < 10^{-6}$) and .275 ($z = 2.84$, $p < 10^{-2}$), respectively. Cross-validation success was greater for the 18-month data than for the artificial data, suggesting that the classification functions for the A, B, C groups are sensitive to nonrandom variance over and above the variance accounted for by the 12-18-month correlations between the variables.

Table 7
Consistency Versus Error In 12-18-Month Stability Data: Cross-Validation
(Data as Decimal Fractions)

	Predicted 18-Month Classification (N=50)			Predicted 18-Month Classification From Error Data (N=500)		
Actual 12-Month Classifi Cation	A	B	C	A	B	C
A	.67	.33	.00	.52	.21	.27
B	.10	.87	.03	.16	.54	.30
C	.09	.27	.64	.24	.21	.55

A χ^2 analysis of goodness of fit indicated that the pattern of cross-validation results from the 18-month data differed significantly from the pattern predicted from the analysis of artificial (error) data, χ^2 (6) - 18.73, $p < .001$. Misclassified subjects from group A were assigned to group B in the actual data; they were equally likely to be assigned to group B or group C in the error data.

Misclassified subjects from group B were most often assigned to group A in actual data; they were most often assigned to group C in the error data. Misclassified subjects in group C were most often assigned to group B in the actual data; they were equally likely to be assigned to group A or group B in the error data. We can confidently reject the hypothesis that unexplained (residual) variance in the 12-18-month correlations in table 5 is entirely unreliable or error variance. In a univariate correlational analysis of temporal stability, this unexplained variance would typically be designated behavioral or measurement "noise." It is apparent, however, that at least part of this variance makes a substantial contribution to the stability of individual differences in patterns of interactive behavior and crying from 12 to 18 months.

Discussion

The analysis of discrete behaviors from Strange Situation data, or from data collected in similar settings, is characteristic of attachment research undertaken from a trait construct perspective (e.g., Coates, Anderson, & Hartup 1972a, 1972b; Feldman & Ingham 1975; Maccoby & Feldman 1972) . Masters and Wellman (1974) have recently reviewed the correlational evidence for stability and inter-correlation among discrete attachment behaviors and have concluded that there is little support for the notion of attachment as a trait construct. The present analysis of discrete-behavior scores from the Strange Situation behavior of a large sample of subjects is consistent with the results reviewed by Masters and Wellman (1974) . There is very little evidence for temporal stability of discrete behaviors, as they have typically been assessed and employed in operational definitions of infant-adult attachment. An analysis of the reliability of individual scores. based on time samples of Strange Situation behavior, however, indicates that neither temporal stability nor significant correlations among discrete attachment behaviors or between these behaviors and external criteria could be expected because typical assessments are too brief to yield reliable data.

In discussing their study of the stability of attachment behaviors from 10-14 and 14-18 months, Coates et al. (1972b) wondered whether the evidence would not have been stronger had they collected longer samples of each subjects' behavior. Indeed, the frequency with which touching, looking, vocalizing, gesturing, approaching, etc., occur in the Strange Situation is quite low (often less than 1.5 per min). A central requirement of time sampling methodology is that samples of a criterion behavior must estimate accurately the parameters of the population from which they are taken (in this case one subject's behavior) in order to be useful. In all behavior sampling techniques, the adequacy of a behavior sample is determined by the interplay of the duration of each sampling interval; the number of times and the rate at, which intervals are sampled; and the duration of each occurrence of the behavior, its rate of occurrence, and its temporal patterning (see Altmann 1974).

Where the behavior in question is as rare as each of the discrete attachment behaviors sampled in the present study, a large number of observation intervals is necessary to obtain reliable estimates of individual scores. Samples of 2, 3, or 5 instances of a behavior easily show fluctuations of 2%--200% on the basis of differences in behavior that should be trivial for the hypothesis in question. Is a child who looks at mother once today and twice tomorrow twice as "strongly attached" in only a day's time?

The present data suggest that the Strange Situation is not the best setting in which to test the hypothesis that discrete behaviors are stable over time and that as indices of attachment they are significantly intercorrelated. At present, this hypothesis remains neither proven, disproven, nor even fairly tested.

What is the likelihood that samples of discrete behaviors based on hours of observation (perhaps even in a variety of settings) would yield valid indices of individual differences in attachment? The chances seem small indeed, for the following reason. When discrete behaviors are used to define attachment operationally, all instances of looking, or vocalizing, or approaching are summed, on the assumption that all instances of phenotypically similar behaviors are equivalent. This assumption is consistent with the a theoretical orientation from which operational definitions often proceed. Unfortunately, this assumption (as a generalization) is manifestly untrue, as many ethological studies of the organization of behavior have demonstrated

(e.g., Baerends 1975) . It is often necessary to take the temporal, situational, and behavioral context of behavior into account in order to distinguish the multiple functions of a given behavior and to derive valid indices of behavior constructs.

All of the discrete behaviors that have been proposed as indices of attachment serve many other behavioral systems as well. Looking, approaching, smiling, or vocalization, for example, are as easily observed in the context of exploration, fear, wariness, affiliation, or play as in the service of attachment. For a total score on a given behavior to serve as an "index" of some attachment construct, there needs to be a consistent influence of attachment across the many instances of the behavior that are combined to produce the total score. If this is not the case, the variance of total scores will not be much influenced by individual differences in attachment. While a measure based on discrete behaviors may look like a useful measure of looking, or distance interaction or attachment, it is more likely that the major consistent influence across unselected instances of a discrete behavior will be the ubiquitous dimensions of temperament.

Regardless of whether a behavior is in the service of exploration, fear, wariness, affiliation, play, or attachment, individual differences in behavioral style can have a consistent (if weak) influence on each instance of the behavior. Unless the relevance of the behaviors scored to attachment is insured by taking context into account, the influence of the attachment behavioral system on total score variance will be less than the influence of activity level, sociability, or emotionality. This is exactly analogous to the situation in which various response biases or response styles are assessed in psychometric research by the construction of scales which have low internal consistency and heterogeneous item content (e.g., Jackson 1970). While the data are not yet in, the prospects for the study of attachment as a trait, even with reliable time sampling data, do not presently seem encouraging (Sroufe & Waters 1977a; Waters, Note 1) .

The major alternative to social learning/ trait models of attachment is the behavioral systems approach developed by Bowlby (1969) and Ainsworth (1972, 1973) and elaborated by Bischof (1975) . The ratings of categories of interactive behavior used in the present study were developed by Ainsworth with specific reference to this theory of the organization of attachment behavior. They explicitly take into account a variety of contextual variables, and, in contrast to typical assessments of discrete attachment behaviors, include assessments of behavior patterns antithetical to the effective functioning of the attachment behavioral system.

The results of the analysis of these categories of interactive behavior and of crying indicate significant stability from 12 to 18 months, especially with regard to behavior toward the mother during reunion episodes. These results are in clear contrast to the results from. the analysis of discrete behaviors.[6] They also contradict the widely held view that behavior is inherently unpredictable and unstable because it is so complexly determined and so sensitive to contextual influences. On the contrary, it seems that the stability of behavior will become apparent *only* when we fully understand its complex determinants and its sensitivity to context.

The impressive stability of profiles or patterns of interactive behavior (table 6) suggests that conventional univariate approaches to continuity in development are not optimal strategies. Indeed, the analysis of the artificial data presented in table 7 suggests that these approaches may be both inefficient and insensitive to important sources of stability in the organization of behavior.

While the present results are encouraging as to certain important characteristics of interactive behavior-assessment scales and classification procedures that can be used in the analysis of Strange Situation data, they are not properly validity data. Despite recent attempts to "validate" the Strange Situation by internal evidence alone (differential response to mother vs. father vs. stranger, etc.), the validity of any procedure as an assessment of an attachment construct depends ultimately upon evidence that the assessment has theoretically relevant patterns a external correlates. Ainsworth et al. (1978) have recently reviewed a wide range of stud-

[6] It has been suggested that the greater stability of the interactive behavior ratings and the A, B, C classifications is due to the fact that these assessments involve fewer categories of (i.e., weaker) prediction than in the time sampling of discrete behaviors. The auxiliary analysis of the ABC categories above indicates that they are not so broad that evidence of stability entails a trivial level of prediction. More important, it has been demonstrated that both the interactive behavior ratings and the A, B, C classifications have a wide range of theoretically significant correlates from early infancy into the third year of life (see Ainsworth et al. [1978] for a review). To date there have been no demonstrations of similar patterns of external correlates for time sampled discrete behaviors. The suggestion that stability is a function of the number of categories employed implies a trade-off between stability and validity. The evidence does not point to this in the present case.

ies establishing the external correlates of interactive behavior categories and A, B, C classifications in the behavior of both mother and infant at home throughout the first year of life, and in a variety of laboratory settings from age 1 year well into the third year of life. These correlates do not follow directly from the evidence for stable individual differences provided in the present study. On the contrary, they are the first steps toward understanding how such stability could have occurred.

The present research was initiated from a perspective within which attachment relationships are assumed to be products of experience, to require environmental support, and to be responsive to environmental changes (Ainsworth 1972; Sroufe & Waters 1977a; Waters, Note 1) . While characteristics of both mother and infant are important in the development of both secure and insecure relationships, the broader social context is important in the stability of the relationship across time. It is predictable that in families under stress early secure attachments may fail. Similarly, at this early stage in development it seems likely that improvements in the family situation could reduce stress on the infant-mother dyad and lead to normative patterns of secure attachment by the middle or end of the second year of life. Evidence that infant-mother relationships can be highly stable from 12 to 18 months in stable family situations provides a useful background against which to study the factors which can lead to change.

Reference Note

1. Waters, E. Traits, relationships, and behavioral systems: the attachment construct and the organization of behavior and development. Manuscript submitted for publication, 1977.

References

Ainsworth, M. Attachment and dependency: a comparison. In J. Gewirtz (Ed.), *Attachment and dependency*. Washington, D.C.: Winston, 1972.

Ainsworth, M. The development of infant mother attachment. In B. Caldwell & H. Riccuiti (Eds.), *Review of child development research.* Vol. 3. Chicago: University of Chicago Press, 1973.

Ainsworth, M.; Bell, S.; & Stayton, D. Individual differences in strange situation behavior of one-year-olds. In H. Schaffer (Ed.), *The origins of human social relations.* London: Academic Press, 1971.

Ainsworth, M.; Blehar, M.; Waters, E.; & Wall, S. *Patterns of attachment.* Hillsdale, N.J.: Erlbaum, 1978.

Ainsworth, M., & Wittig, B. Attachment and exploratory behavior of one-year-olds in a strange situation. In B. Foss (Ed.), *Determinants of infant behavior. Vol. 4.* New York: Barnes & Noble, 1969.

Altmann, J. Observational study of behavior: sampling methods. *Behaviour,* 1974, *49,* 227-267.

Baerends, C. A model of the functional organization of incubation behavior. In C. Baerends & R. Drent (Eds.), The herring gull and its egg. *Behaviour Supplement,* 1975, 17, 261210.

Bischof, N. A systems approach to the functional connections of fear and attachment. *Child Development, 1975, 46, 801-817.*

Bowlby, J. *Attachment and loss. Vol. 1. Attachment.* New York: Basic, *1969.*

Coates, B.; Anderson, E.; & Hartup, W. Interrelations in the attachment behavior of human infants. *Developmental Psychology, 1972, 6, 218-230.* (a)

Coates, B.; Anderson, E.; & Hartup, W. The stability of attachment behaviors in the human infant. *Developmental Psychology, 1972, 6, 231-237.* (b)

Cohen, J. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 1960, 20, 37-46.*

Cronbach, L. Coefficient alpha and the internal structure of tests. *Psychometrika, 1951, 16, 297-334.*

Cronbach, L., & Meehl, P. Construct validity in psychological tests. *Psychological Bulletin, 1955, 52, 281-302.*

Feldman, S., & Ingham, M. Attachment behavior: a validation study in two age groups. *Child Development, 1975, 46, 31930.*

Fliess, J.; Cohen, J.; & Everitt, B. Large sample standard errors of kappa and weighted kappa. *Psychological Bulletin, 1969, 72, 323-327.*

Jackson, D. A sequential system for personality scale development. In C. Spielberger (Ed.), *Current topics in clinical and community psychology. Vol 2.* New York: Academic Press, *1970.*

Maccoby, E., & Feldman, S. Mother-attachment and stranger-reactions in the third year of life. *Monographs of the Society for Research in Child Development, 1972, 37(1,* Serial No. *146).*

Masters, J., & Wellman, H. Human infant attachment: a procedural critique. *Psychological Bulletin, 1974, 81, 218-237.*

Nunnally, J. *Psychometric theory.* New York: McGraw-Hill, *1967.*

Sroufe, L., & Waters, E. Attachment as an organizational construct. *Child Development, 1977, 48, 1184-1199.* (a)

Sroufe, L., & Waters, E. Heart rate as a convergent measure in clinical and developmental research. *Merrill-Palmer Quarterly, 1977, 23, 3-27.* (b)

Wiggins, J. *Personality and prediction: principles of personality assessment.* Reading, Mass.: Addison-Wesley, *1973.*