Personal Relationships, **2** (1995), 247–261. Printed in the United States of America.
 1350–4126/95 $7.50 + .10

On the instability of attachment style ratings

MARK W. BALDWIN AND BEVERLEY FEHR
University of Winnipeg

Abstract

We examined the stability of ratings on the Hazan and Shaver (1987) single-item attachment style scale in a number of data sets, gathered by us and other researchers. Approximately 30% of subjects overall changed their attachment style classifications over a relatively short time span (ranging from 1 week to several months). The highest rate of instability was observed in subjects who classified themselves as anxious-ambivalent—the majority of whom changed their ratings from one time to the next. Given these findings, we explore the methodological and conceptual implications of instability in attachment style ratings. With regard to the former, we question the current practice of selecting subjects for participation in research based on responses to the attachment style questionnaire administered on a different occasion. Our findings suggest that a substantial proportion would change their style rating in the interim. In terms of conceptualization, we examine a number of different explanations for the observed instability and propose that it may reflect variability in the underlying construct, rather than a lack of continuity in style or unreliability of measurement. From this perspective, an individual's response to an attachment style questionnaire reflects the relational schema that is activated at that moment, rather than an enduring general disposition or trait. Stability in ratings is therefore neither assumed nor expected.

Bowlby (1969, 1973, 1980) held that infants develop expectations about the availability and responsiveness of their caregivers, and that these working models form an emotional bedrock for their developing personality and interpersonal relations. He proposed that these expectations are largely established by age 5, although they may be modified and elaborated somewhat during childhood and early adolescence. Bowlby took the position that "whatever expectations are developed during those years tend to persist relatively unchanged throughout the rest of life" (Bowlby, 1973, p. 202). On the basis of this claim of temporal continuity, researchers have begun to study how attachment styles (Ainsworth, Blehar, Waters, & Wall, 1978) might shape adults' significant relationships and personalities. Much of the research has followed from Hazan and Shaver's (1987) pioneering work, in which adults were asked to identify their own attachment style on a single-item measure. For this measure, subjects select from paragraphs describing the three styles: secure (i.e., comfortable with dependency and closeness to others), avoidant (i.e., uncomfortable with others' desires for closeness and dependency), or anxious-ambivalent (i.e., desiring a high level of closeness to others, but anxious that others might not want to be close).

This research was supported by Social Sciences and Humanities Research Council of Canada grants held by each author. We would like to thank Patrick Keelan, Carole Pistole, Marilyn Senchak, Gil Becker, Kim Bartholomew, and Elaine Scharfe for their helpful comments on an earlier draft of this article, without meaning to imply their agreement with the ideas we present here.

Address correspondence to Mark Baldwin or Beverley Fehr, Department of Psychology, University of Winnipeg, Winnipeg, Manitoba, R3B 2E9 Canada. E-mail may be sent to Mark Baldwin at baldwin@uwinipeg.ca or to Beverley Fehr at uowbaf@ccm.umanitoba.ca.

While this scale has received some criticism because it is a single-item, categorical measure, it has nevertheless enjoyed widespread use (e.g., Feeney & Noller, 1990, 1991; Hazan & Shaver, 1990; Kobak & Hazan, 1991; Mikulincer, Florian, & Tolmacz, 1990; Mikulincer & Nachshon, 1991; Pistole, 1989). Some researchers have begun to complement this instrument with other measures based on continuous ratings (e.g., Brennan & Morris, 1993; Brennan, Shaver, & Tobey, 1991; Keelan, Dion, & Dion, 1994; Shaver & Brennan, 1992), but many others continue to use or report the categorical measure alone (e.g., Eiden, Leonard, & Senchak, 1993; Kirkpatrick & Hazan, 1994; Kirkpatrick & Shaver, 1992; Pietromonaco & Carnelley, 1994; Sprecher et al., 1994; Tidwell, Cooper, & Shaver, 1993), presumably because of its ease of administration and wide acceptance in the literature.

As implied by the term "style," most researchers have taken an individual difference approach to studying attachment behavior. The tacit assumption has been that attachment style is a trait-like construct, which remains relatively invariant over time and across different close relationships. Hazan and Shaver's attachment style scale has been generally regarded as an adequately reliable measure of this stable trait. This assumption presumably underlies such research practices as preselecting subjects based on their responses to the measure administered previously (e.g., mass testing of subject pools at the beginning of the academic year) or categorizing subjects into different styles based on their responses to the attachment style scale completed weeks earlier or later.

There is, however, a growing uneasiness in the literature about both methodological and conceptual issues, as reflected in many of the commentaries on Hazan and Shaver's (1994a) recent overview of adult attachment theory. As part of this debate, we believe that it is important to examine the popular three-category measure, particularly with regard to issues of stability and reliability. In so doing, it is necessary to distinguish among a number of related concepts. We will use the term "(dis)continuity" to refer to claims that the attachment style laid down in the early years persists, relatively unaltered, across the life span. We will use the term "(in)stability" to refer to the extent to which subjects' self-classification into one of the three styles changes from one occasion to the next. Finally, we will use the term "(un)reliability" to describe that portion of instability in scale scores attributable to inadequate measurement. These distinctions are important because instability (changes in subjects' attachment style ratings) could result from either unreliability or discontinuity. In other words, continuity in an underlying construct might, in fact, exist, even though a scale designed to measure that construct lacked the psychometric property of reliability. Conversely, a scale might be accused of being unreliable when instability in scale scores actually veridically mirrored instability in the underlying construct.

Indirect evidence for the stability and reliability of the measure has been provided in studies that show the many correlates of self-rated attachment styles (see, e.g., Shaver & Hazan, 1993, for a review). Little direct evidence has been published, however, to document either the stability of adult attachment styles or the reliability of the Hazan and Shaver scale. With regard to the latter, a single-item measure obviously does not lend itself to standard reliability assessments in terms of internal consistency, split-half reliability, item–whole correlations, and so on. However, test–retest reliability, or consistency in responses over time, can be assessed. Pistole's (1989) study is most commonly cited as evidence of the reliability of the Hazan and Shaver scale; she reported "adequate consistency" in their measure over a 1-week period (p. 507). Other reports that have appeared in the literature more recently will be discussed shortly.

We first became interested in the stability of attachment styles when, inspired by the burgeoning literature, we designed

a number of studies to explore further the correlates and underlying dynamics of attachment styles. In most populations, the styles are observed in unequal proportions, with approximately 55% secure, 25% avoidant, and 20% anxious-ambivalent subjects (Hazan & Shaver, 1994a). For several of our studies, we wished to recruit an equal number of each style, and so we included the Hazan and Shaver (1987) measure in a mass-testing package that was administered at the beginning of the academic year. At the end of one study (Baldwin, Fehr, Keedian, Seidel, & Thomson, 1993), we had subjects complete the attachment style questionnaire again as a reliability check. To our surprise, very few subjects who had rated themselves as anxious-ambivalent during the initial mass testing session now endorsed that description; most of these subjects rated themselves as secure. This finding was unsettling to us, given that we had attempted to select equal numbers of the three styles based on their Time 1 (mass testing) data. When we compared the ratings for the other two styles, the results were more encouraging, although we still encountered a remarkable degree of instability. It was possible, of course, that the instability we observed was peculiar to this sample or this particular experimental context. However, it also was possible that the construct and/or its measurement might not be as stable as we had assumed.

Fortunately, test–retest data were available for six of our attachment style studies; these data allowed us to determine the pervasiveness of instability in attachment style ratings among University of Winnipeg students. We also wanted to examine stability ratings in samples other than our own. Thus, we solicited data from researchers who had also measured attachment style on more than one occasion.[1]

1. We would like to thank Kelly Brennan, Patrick Keelan, Carole Pistole, and Marilyn Senchak for so generously supplying us with their data. Their cooperation is greatly appreciated.

Method and Results

Our studies

Data were drawn from six studies that had been conducted over a two-year period. All subjects were undergraduate students from the introductory psychology subject pool at the University of Winnipeg. Table 1 lists the subject composition and test–retest intervals of these studies. With all of these smaller samples combined, the entire sample consisted of 159 female and 62 male subjects.[2] Their average age was 20.5 years.

All subjects completed the Hazan and Shaver single-item measure during an initial mass testing session and again 3 to 4 months later, depending on the study. Several features of the attachment style data (aggregated across studies and displayed in Table 2) are worth noting. First, the marginal proportions at Time 1 (mass-testing pretest) and Time 2 (subsequent experimental session) are not only very similar, but are also generally consistent with the typical proportions in the literature (and with the proportions in our entire introductory psychology subject pool [N=689], which were 56% secure, 32% avoidant, and 12% anxious-ambivalent). On the basis of this finding, one might be tempted to conclude that the attachment style ratings were, in fact, stable. Further inspection of the contingency table, however, shows that such a conclusion would not be justified. It was only the proportions, and not the subjects on which they were based, that stayed the same.

The contingency data can be analyzed and interpreted in a number of ways, some of which are more informative than others.[3] Probably the most telling analysis is

2. Some subjects participated in more than one of the post-studies. Their style rating from the first post-study in which they participated was used in all analyses.

3. Analysis of the frequencies yielded a highly significant chi-square, $\chi^2(4) = 65.41$, $p > .001$. The pre- and post-ratings are thus statistically related; how-

a simple examination of the proportions. Of the individuals who were secure at Time 1, 19.5% (24/123) changed their self-classifications on their Time 2 rating. Of the individuals who were avoidant, 42.5% (31/73) changed their self-rating. Finally, of the individuals who described themselves as anxious-ambivalent, fully 68% (17/25) claimed to be a different style at Time 2. Collapsing across the three styles, the overall proportion changing was 32.6% (72/221). Looked at differently, the proportion of agreement between Time 1 and Time 2 was only .674 (roughly two-thirds), when .446 would be expected on the basis of chance alone, given the marginal probabilities (e.g., Cohen, 1960). One way of evaluating the proportion of agreement, while adjusting for chance, is with the statistic kappa (Cohen, 1960), which had a value of .41 in this sample. Cicchetti and Sparrow (1981) provided guidelines for interpreting kappa: poor agreement as ≤ .40; fair agreement as .40–.59; good as .60–.74; and excellent as .75–1.00. According to these guidelines, the value in the present sample of .41 just passed the criterion for a fair level of agreement.

The proportion of the sample changing was generally consistent across genders (women, 31.0%; men, 35.5%) and across the six smaller studies from which we drew our sample (with values ranging from 16.7% to 56.3%; see Table 1). In fact, it seems that the most stable aspect of these data was the presence of instability.

ever, these findings do not provide a strong recommendation in the absence of an interpretable index of agreement (Cohen, 1960). A correlation analysis of pre- and post-classifications (transformed to dummy variables) showed that being secure at Time 1 was moderately correlated with being secure at Time 2 ($r = .50$); for avoidants the correlation was .41, and for anxious-ambivalents it was only .24 (all $ps < .001$). While both the chi-square and correlational analyses are significant, the magnitude of the correlations, especially for the avoidant and anxious groups, is hardly convincing evidence of stability.

Data from other researchers

Next, we wanted to ensure that the instability in attachment style ratings we observed was not unique to our sample. We searched for other studies in which attachment style had been measured more than once, and we were provided test–retest data by Pistole (1989), Keelan et al. (1994), and Brennan and Shaver (1990; also referred to in Shaver & Brennan, 1992). These data are summarized in Table 1. The test–retest intervals ranged from 1 week to 12 months, and the overall rates of instability ranged from 19.8% to 28.9%. It is particularly worth noting that the rate of instability for those endorsing the anxious-ambivalent description at Time 1 ranged from a low of 42.4% to a high of 75%.

Data from all samples combined

The data collected by other researchers showed degrees of instability comparable to ours, although the exact rate of change was not identical across studies. The accuracy of the estimated rate of change varied according to the sample size, of course; note that the highest (56.3%) and lowest (16.7%) estimates were found in samples of ours with fewer than 20 subjects. To provide a better estimate of the degree of stability/instability, we combined all of the data sets. The pooled data are displayed in Table 3. Note, first, that the proportions of secure, avoidant, and anxious-ambivalent subjects both at Time 1 and Time 2 are consistent with the literature. Once again, however, the stability in proportions at the group level conceals a notable degree of instability at the individual level. The overall rate of change from Time 1 to Time 2 for this combined sample was 28.0% (145/517; kappa = .51). Further, the median rate of change for the nine samples was 28.9% (see Table 1), and the unweighted mean was 30.0%. Returning to the pooled data, of the subjects who reported themselves secure at Time 1, 17.2% changed at Time 2; of avoidants, 33.5%

Table 1. *Comparison of studies: Proportions of subjects changing attachment style ratings*

Sample	*N* (female, male)	Approximate Time Lag (wks)	Degree of Instability (% change)			
			Time 1 (Secure)	Time 1 (Avoidant)	Time 1 (Anxious-Ambivalent)	Total Sample
Baldwin et al., 1993	16 (11, 5)	16	16.7	80.0	80.0	56.3
Research Seminar Study 1	65 (40, 25)	16	12.5	41.3	66.7	35.4
Research Seminar Study 2	51 (37, 14)	16	11.8	50.0	80.0	27.5
Research Seminar Study 3	31 (19, 12)	16	28.6	33.3	0.0	29.0
Tinio, 1992	12 (6, 6)	16	11.1	33.3	n/a[a]	16.7
Barnes, 1991	46 (46, 0)	12	31.0	33.3	50.0	32.6
Six samples pooled	221 (159, 62)		19.5	42.5	68.0	32.6
Pistole, 1989	67 (42, 25)	1	15.6	21.5	75.0	23.9
Keelan, Dion, & Dion, 1994	101 (69, 32)	16	12.1	20.7	50.0	19.8
Shaver & Brennan, 1992	128[b]	40	18.9	31.0	42.4	28.9
Nine samples pooled	517		17.2	33.5	55.0	28.0
Senchak & Leonard, 1992	718(363, 355)	52	13.0	69.0	88.5	25.8

[a]No subject in this sample rated self as anxious/ambivalent at Time 1.
[b]We did not receive gender information for this sample.

Table 2. *Our six studies combined: Frequencies of subjects endorsing attachment styles at Time 1 and Time 2.*

	Style at Time 2			
Style at Time 1	Secure	Avoidant	Anxious-Ambivalent	Marginal (%)
Secure	**99**	17	7	55.7
Avoidant	22	**42**	9	33.0
Anxious-Ambivalent	8	9	**8**	11.3
Marginal (%)	58.4	30.8	10.9	

Table 3. *Pooled data from nine samples: Frequencies of subjects endorsing attachment styles at Time 1 and Time 2.*

	Style at Time 2			
Style at Time 1	Secure	Avoidant	Anxious-Ambivalent	Marginal (%)
Secure	**231**	26	22	54.0
Avoidant	43	**105**	10	30.6
Anxious-Ambivalent	28	16	**36**	15.5
Marginal(%)	58.4	28.4	13.2	

changed; and of anxious-ambivalents, 55.0% changed their style rating.

Corroborating evidence from another sample

We also received data from Senchak and Leonard (1992b; see also Senchak & Leonard, 1992a), which were based on a sample of 363 wives and 355 husbands, assessed first when they applied for their marriage license and then again approximately 1 year after their wedding. This sample was unusual in several respects (e.g., couples going through the major life transition of marriage, which might be expected to make them somewhat more secure; see Kobak & Hazan, 1991), which was the reason we chose to analyze it separately. In fact, approximately 80% rated themselves as secure, which is a much higher proportion than is usually found in the general population. However, even with the distinctiveness of this sample, the degree of instability in the attachment style ratings was of the same order of magnitude as that found for the other samples combined, with an overall rate of change of 25.8% (kappa = .20)(see Table 1).[4]

Additional analyses

One might suggest that the observed instability is simply due to measurement error inherent in the categorical response format. One way to examine this question is by comparing the measure with continuous indicators of attachment styles, which some

4. It is worth noting that the rate of instability was fairly consistent across studies, even though the exact form of the question varied in terms of whether subjects were asked to choose the description that "best describes your feelings" (Hazan & Shaver, 1987) in close relationships, romantic relationships, or close romantic relationships.

researchers have begun to use in hopes of increasing reliability. Although continuous Time 1 measures were not available from all of our subjects, continuous Time 2 measures (on which they rated each style paragraph on a 7-point scale ranging from "does not describe me at all" to "describes me very well") were available for 171 subjects. If the variability in the Hazan and Shaver measure was simply due to measurement error produced by the categorical format, the categorical ratings should not correlate particularly well with the continuous ratings. To the contrary, when a discriminant analysis was performed using the continuous ratings to predict *concurrent* categorizations of attachment style, 96.5% of subjects were correctly classified (97% of secures, 96.2% of avoidants, 94.4% of anxious-ambivalents). Thus, these two measurement approaches gave highly similar results when they were administered in the same session. However, when a second discriminant analysis was performed using the continuous ratings at Time 2 to predict self-categorizations at Time 1, this analysis correctly classified only 62.6% of subjects (75.8% of secures, 47.4% of avoidants, 47.8% of anxious-ambivalents). These results indicate a high correspondence between measures of attachment style when they are administered concurrently; there is much lower correspondence when the measures are administered at different times.

Finally, it seems reasonable to suspect that the people who change their style ratings might be those who also experience a change in their relationship status. In fact, some recent studies have supported this possibility (see, e.g., Feeney, Noller, & Callan, 1994; Hammond & Fletcher, 1991; Kirkpatrick & Hazan, 1994; Senchak & Leonard, 1992a). Although we did not have detailed information on the course of our subjects' relationship histories, relationship status information at both measurement times was available for 175 of the subjects, which allowed us to make a rough assessment of this hypothesis. However, when change in ratings of relationship status was correlated with change in attachment style ratings, there was no evidence supportive of the hypothesis (see also Scharfe & Bartholomew, 1994).

Recent reports from other researchers

Since we began to collect and analyze these data, we have become aware of a few recent reports in which the rates of change in attachment measures are mentioned. Citing an unpublished manuscript by Hazan, Hutt, and Markus (1991), for example, Shaver and Hazan (1993) reported a 22% rate of change. More recently, Kirkpatrick and Hazan (1994) collected test–retest data (N=172) over a 4-year time lag. Their results were very similar to ours: the overall rate of change was 30%, with 17% of secures, 39% of avoidants, and 50% of anxious-ambivalents changing (kappa = .51).

Finally, Scharfe and Bartholomew (1994) conducted an extensive analysis of 8-month stability in the Bartholomew and Horowitz (1991) four-category attachment measure. This was an important addition to the debate on measurement, as instability in the Hazan and Shaver measure might simply be the result of trying to force a four-category construct into the procrustean bed of a three-category measure. Scharfe and Bartholomew selected only subjects who were involved in stable relationships (of 2 years minimum), which might be expected to make their data more stable than those in other studies. Nevertheless, they reported rates of change of 37% (kappa = .42) for female and 44% (kappa = .26) for male subjects; these rates showed even more instability than was observed in the three-category data. Of the subjects who endorsed one of the three insecure styles at Time 1, 53% (35/66; see their Table 1, p. 32) switched to a different style at Time 2. These findings are generally consistent with our results.

Scharfe and Bartholomew were considerably more sanguine than we are about the stability of attachment styles. They found, for example, that when they used an interview procedure (involving hour-long struc-

tured interviews, coded by judges who had completed a 200-hour training program), test–retest instability dropped to 23% (27% for Time 1 insecures). They also demonstrated that when test–retest correlations for continuous measures were disattenuated, using estimates of internal consistency, the adjusted estimates of "true stability" moved up into an acceptable range (see also Feeney, Noller, & Callan, 1994). We agree with them that interview and continuous measures are surely superior in many ways to 1-item self-reports, and that increasing the internal consistency of measures should increase test–retest correlations. Nonetheless, the issues of variability persist. It is proving remarkably difficult to generate the sought-after measures of high internal consistency, and even when individuals in highly stable relationships are assessed using commendably ambitious interview approaches, considerable attachment instability remains. Add to this the finding that interview, self-report, and third-party attachment ratings often show surprisingly little overlap (e.g., Bartholomew & Horowitz, 1991, p. 231, fn. 1), and one is left with a sense that serious issues of measurement and conceptualization are still unresolved (see also Bartholomew, 1994; Noller & Feeney, 1994).

Discussion

The issue identified across the data sets examined is that approximately 30% of people changed their self-rated attachment style over a period of months. This rate of change seems problematic. It is especially problematic in the case of anxious-ambivalent subjects—the majority of whom changed their style, even over a 1-week period (Pistole, 1989). This raises at least two issues of relevance to adult-attachment researchers, one methodological and one conceptual.

Implications for research practice

At a very practical level, the instability we found has implications for how research is conducted in this area. Researchers sometimes administer attachment measures during mass-testing sessions (e.g., Pietromonaco & Carnelley, 1994; Shaver & Brennan, 1992) or else classify subjects into different styles based on responses provided a few weeks or months before or after the experimental session (e.g., Eiden et al., 1993; Feeney & Noller, 1991; Hazan & Shaver 1990, Study 2; Kirkpatrick & Davis, 1994; Kirkpatrick & Shaver, 1992; Kobak & Hazan, 1991; Mikulincer et al., 1990). Our findings suggest that nearly one-third of these subjects would not have endorsed the same style at the time of the experiment. In fact, the *majority* of subjects whom the researchers regarded as anxious-ambivalent probably would have classified themselves as secure or avoidant if they had completed the scale during the experimental session.[5]

In light of this finding, it becomes understandable that the results are often weaker or less interpretable in such studies than in those where the attachment style scale is administered concurrently with the other measures. For example, Mikulincer and Nachshon (1991) described three studies that examined the relation between attachment styles and patterns of self-disclosure. In the first and third studies, subjects completed all of the measures in the same session. In the second study, subjects completed the attachment questionnaire and later were telephoned and asked to partici-

5. We do not wish to over-interpret the finding that the highest rate of change was for people who characterized themselves as anxious-ambivalent at Time 1. This finding may reflect the psychological nature of this attachment orientation, as ambivalence implies a conflict among alternative models of relatedness. On the other hand, even if the psychological likelihood of change were exactly equal across the styles, one would still predict greater rates of change for anxious-ambivalents simply because the base rates for this style are low (Scharfe & Bartholomew, 1994). Irrespective of the cause of this finding, the methodological point is clear: Conclusions about any correlates of the anxious-ambivalent style are extremely suspect if they are based on measures taken at a different time, given that over half of these subjects would endorse a different style if asked concurrently.

pate in the self-disclosure study (the time lapse was not specified). Interestingly, it was in this study that the authors reported: "Findings supported our predictions for secure and avoidant persons, but were at odds with our predictions for ambivalent persons" (p. 326).

In a recent article involving a longitudinal study of dating couples, Kirkpatrick and Davis (1994) correlated subjects' initial attachment style ratings with concurrent and subsequent relationship outcomes. At the first follow-up (approximately 1 year later), the authors obtained a number of findings described as "paradoxical" (p. 508) or otherwise difficult to interpret.

These kinds of findings are not uncommon, as unpredicted findings frequently emerge when attachment style is assessed at a different time from the other measures (see, e.g., Collins & Read, 1990; Feeney and Noller, 1991; Kirkpatrick and Shaver, 1992). Most of these authors offered explanations for their findings, and we do not wish to dispute their interpretations; indeed, any data set might be expected to include some anomalous findings. However, we would like to raise the possibility that some of the unexpected, surprising, and paradoxical findings in the literature might be attributable to shifts in subjects' attachment ratings from one measurement to the next (as also acknowledged by Kirkpatrick & Davis, 1994, p. 510). Our instability findings suggest that if subjects do not classify themselves during the experimental session, investigators should be aware that even 1 week later a substantial proportion of them, especially those who endorse the anxious-ambivalent style, could choose a different style rating. This obviously has considerable implications for the designing of future studies and for the interpretation of previously published results.

Instability of Attachment Style Ratings: Conceptual Issues

Our main purpose in gathering and reporting these data was to document the instability of attachment style ratings and to alert adult-attachment researchers to the methodological and conceptual implications of instability in the categorical measure. The studies reviewed here were not specifically designed to provide critical tests among competing hypotheses for exactly why this instability is observed. Nonetheless, it seems appropriate to consider a number of alternatives that could guide future research on the issue.

Lack of continuity in attachment style?

The instability in attachment style ratings we observed may indicate less long-term continuity in attachment style than has generally been supposed. As Hendrick and Hendrick (1994) noted, many observers find it difficult to believe that "The way the infant attaches at 1 year of age mostly determines the way the adult attaches at age 21" (p. 39). Although, as Hazan and Shaver (1994b) point out, this is a somewhat misleading caricature of the claims of attachment theory, the theory does propose a significant degree of continuity, particularly from adolescence onward (Hazan & Shaver, 1994b). It would be premature to draw strong conclusions about continuity from the present data; however, one relevant observation can be made. When one examines the various data sets (see Table 2), it becomes apparent that the degree of instability is not obviously related to the amount of time that elapsed.

For example, Pistole's 1-week interval yielded instability rates roughly equivalent to Brennan and Shaver's 10-month interval and even Kirkpatrick and Hazan's (1994) 51-month interval (where they reported a 30% rate of change). For the nine studies considered here, the correlation between the length of time between measurements and the degree of change was near zero (r =.06). Similarly, Bartholomew (1993) reported that the magnitude of test–retest correlations using continuous attachment style scales does not vary with the time interval between testing sessions. If attach-

ment style ratings do not become systematically more unstable as the amount of time elapses, this suggests that the instability we found may be due to unreliability in attachment style measures (see also Griffin & Bartholomew, 1994; Scharfe & Bartholomew, 1994).

Unreliability of measures?

In the data sets we examined, the Hazan and Shaver categorical measure did not fare particularly well in terms of test–retest reliability. Perhaps the reason is that a categorical approach is inherently unstable, as some individuals might not fit into any one category exactly. One could attempt to design a more reliable assessment instrument by adding more questions, using continuous ratings (e.g., Collins & Read, 1990; Feeney, Noller, & Hanrahan, 1994; Simpson, 1990; West, Sheldon & Reiffer, 1987), or adopting an interview approach (e.g., Bartholomew & Horowitz, 1991; Main & Goldwyn, 1988). These and other strategies have been suggested by a number of researchers, including Hazan and Shaver (1987), and they seem appropriate avenues to explore. In studies where the subjects have given continuous ratings of the three attachment style descriptions or of the individual statements from the descriptions, test–retest correlations have been in the .60 range (e.g., Collins & Read, 1990; Feeney, Noller, & Callan, 1994; Hammond & Fletcher, 1991; Keelan et al., 1994; Levy & Davis, 1988; Shaver & Brennan, 1992). Thus, test–retest reliability is improved by the use of such measures; however, the correlations still tend to be moderate. For a core, influential aspect of personality, attachment style has proven remarkably difficult to measure.

From our own data, one finding argues against a simple finger-pointing response in which the categorical approach is blamed for the observed instability. When the categorical measure was compared with continuous measures of the three styles, which are often considered more meaningful, discriminant analyses showed 96.5% agreement when the measures were taken concurrently. It is only when the continuous measures were compared with the categorical measure taken at a different time that low correspondence once again was observed (i.e., 62.6% agreement). These data, along with the typically rather low test–retest correlations for the continuous measures, suggest that, whereas both measurement approaches may suffer from response biases and other sources of error, there is no compelling reason to conclude that the categorical measure is markedly less reliable than continuous measures.

When considering the categorical measure, it is not clear what one should deem an acceptable level of reliability. Optimists might argue that we should be pointing out that the glass is two-thirds full, rather than one-third empty. Kirkpatrick and Hazan (1994), for example, considered a 70% stability rate "remarkably high" (p. 135). However, the fact that, across all of the samples, nearly one-third of subjects overall (and the majority of anxious-ambivalents) changed their style rating does not seem very encouraging, especially if one assumes that the underlying construct is invariant and continuous over time. We believe the data call into question that assumption, and we wish to suggest an alternative, but nevertheless quite optimistic, interpretation.

Variability in the underlying construct?

The interpretation of instability in a measure depends entirely on how stable the underlying variable is presumed to be (Nunnally, 1978). If the underlying variable is presumed to be invariant, instability implies an unreliable measure. If, on the other hand, one allows for variability in the attachment style construct, then asking for a demonstration of test–retest reliability in an attachment style scale may be making an inappropriate request. This request would be akin to asking for the reliability of mood scales, for example, where test–retest reliability is not expected because it is assumed that the underlying construct is constantly

changing. What is expected of such measures is that they demonstrate adequate validity (see Russell, 1989), as the attachment style scale has done. As mentioned earlier, the Hazan and Shaver measure has shown a meaningful pattern of relations with both relationship and nonrelationship variables, particularly in studies where the measures have been administered concurrently.

We wish to present the view that the attachment style construct, rather than just its measurement, warrants reexamination. We start with the hypothesis that instability in self-reported attachment style stems not from unreliability of measurement, but rather from psychologically meaningful variability. If so, what kind of theory would be required to account for this variability? Could this approach also accommodate various findings that are somewhat awkward for an individual differences formulation? Consider, for example, that in the infant-attachment literature it is well established that children often exhibit a different "attachment style" toward their mother than toward their father, when assessed using the Strange Situation procedure (see Bretherton, 1985). We (Baldwin, Keelan, Fehr, & Koh-Rangarajoo, 1994) recently found that adults also self-report different "styles" in different significant relationships. Consider also the finding reported earlier that significant life experiences, such as marriage, may shift an adult's attachment style. This finding obviously raises questions about the assumption that styles derive directly from childhood experiences with caregivers; after all, the person whose style has changed still has the same childhood as before. What might change, however, is which memories come to mind, or how the person interprets those events (e.g., Main, Kaplan, & Cassidy, 1985; see also Blatt & Homann, 1992, for a similar analysis of findings in the depression literature).

For attachment theory to integrate findings of meaningful variability there will need to be a shift in emphasis away from a trait-based, individual differences approach, to a more thoroughgoing social–cognitive conceptualization (similar to that proposed by various personality theorists, e.g., Cantor & Kihlstrom, 1987; Mischel, 1973). Baldwin (1992) observed that a number of researchers, across a wide range of literatures, are beginning to study relational schemas, or "cognitive structures representing regularities in patterns of interpersonal relatedness" (p. 461). A relational schema is assumed to comprise a self-schema, a schema for another person, and a script for the kinds of interactions that typically occur between self and other. Thus, many writers are pursuing the same type of analysis as Bowlby proposed in his discussion of working models (for examples in the infant-attachment literature, see Bretherton, 1985, 1990; Crittenden, 1990; Sroufe & Fleeson, 1985).

In particular, one idea that is more prevalent outside the adult-attachment literature is that people can have relational schemas, or working models, of many different forms of interpersonal relatedness. A person might have a schema for "I am anxious and she comforts me," for example, as well as one for "I trust her and she hurts me" and one for "I reach out for her and she does not understand." In other words, the same person quite easily could have schemas for the various types of relational expectations that are presumed to underlie all three attachment styles. These expectations may have developed in different contexts, with various significant others, or at different times (e.g., Main et al., 1985).

An individual's momentary attachment orientation, then, would derive from the subset of memories, self-concepts, and interpersonal expectations activated at the time. Conversely, the degree of stability in attachment styles would reflect the stability in which schema was activated. From this perspective, research into attachment orientations would involve delineating the factors contributing to stability and fluctuation in activation patterns. For example, a person's attachment orientation might fluctuate on a momentary basis, as a result of interpersonal contexts or subtle cues that activate one schema or another. Some re-

search (e.g., Baldwin, Carrell, & Lopez, 1990; Baldwin & Holmes, 1987; Higgins, Bond, Klein, & Strauman, 1986) has shown that self-evaluative relational schemas can be primed by situational factors; the same principle may apply to the working models involved in attachment behavior. A person may be more secure after watching a heart-warming Walt Disney motion picture, for example, or more avoidant after watching the movie *Fatal Attraction*. These shifts in "state of mind" (e.g., Horowitz, 1979) may represent far more than simple changes in mood. They may bring with them different constellations or networks of memories, expectations, interpretations, and behavioral tendencies (Baldwin et al., 1994).

Second, as other studies have revealed, people's active attachment schemas probably reflect the state of their current significant relationships, such as whether or not they are happily married or involved in a dating relationship (e.g., Feeney, Noller, & Callan, 1994; Hammond & Fletcher, 1991; Kirkpatrick & Hazan, 1994; Kobak & Hazan, 1991; Senchak & Leonard, 1992a). In our data, there was no indication that changes in dating status predicted changes in attachment style ratings, but the measures were fairly crude and not specifically designed to test this hypothesis. In the developmental literature, however, there is an indication that stability in an infant's attachment behavior mostly reflects stability in the primary relationship, rather than some characteristic internal to the child. In an article reviewing the literature on the Strange Situation procedure, Lamb, Thompson, Gardner, Charnov, and Estes (1984) discussed a number of studies reporting rates of change in infants' attachment styles in the range of 40%. Much of this change was attributable to changes in the life circumstances of the mothers and children: "the Strange Situation assessments may indeed reflect the *current* but not necessarily *enduring* status of mother–infant interaction" (Lamb et al., 1984, p. 136).

Finally, there probably is some stability that derives from an individual's preferred, or chronically most accessible, relational schema. It would be folly to deny that people have stylistic preferences in their processing of social information and in the interpretive schemes they employ for disambiguating the nuances of interpersonal experience. In the current sample, for instance, approximately two-thirds of subjects did give the same self-categorization both times. The present social–cognitive view of individual differences is quite different, though, from others that leave little room for meaningful intra-individual variability in information processing. Even though an individual may have one relational schema that tends to be chronically accessible, he or she might have a wider repertoire of other schemas that can be activated by specific relationship partners, current contexts and goals, and so on.

When placed in the framework of current social–cognitive theory, instability in the Hazan and Shaver (1987) attachment style measure does not seem particularly problematic to us. We intend to continue using this measure, along with its continuous counterparts, on the assumption that they measure a current attachment orientation, based on activated relational schemas, rather than with the view that they assess an invariant, quasi-immutable personality trait that was laid down in childhood. The use of these measures is problematic only when researchers classify subjects based on self-ratings taken at a different time.

If attachment researchers are going to continue to talk about and to work on developing reliable measures of chronic attachment "style," we suggest that the issues of variability we have raised here must be considered. What does it mean to talk of a single "attachment style," for example, if people often endorse different attachment descriptions for different significant relationships and even for different romantic relationships (e.g., Baldwin et al., 1994; see also Kobak, 1994; Lewis, 1994, among others, for similar arguments)? As we have discovered ever since we began exploring a social–cognitive analysis (e.g., Baldwin,

1992; Baldwin et al., 1993), once one begins to examine working models in detail, one inevitably confronts issues of variability and instability. Indeed, even Shaver, Collins, and Clark (in press) stated that "Just as it is incorrect to speak of a single model of self or others, it may be incorrect to speak of a person's single attachment style." We could not agree more.

In conclusion, the popular Hazan and Shaver (1987) single-item measure of attachment styles shows a notable degree of instability that can be accounted for in a number of ways. We considered three explanations: a lack of long-term continuity in people's attachment behavior, a lack of reliability in the measurement of truly stable attachment styles, or the presence of short-term instability in individuals' "states of mind" with respect to attachment. We tend to favor the latter view, as it is most congruent with current thinking on cognitive structures representing interpersonal experience. Moreover, data are beginning to accumulate to attest to the value of an explicitly social–cognitive view of attachment by delineating the interpersonal expectations that contribute to this domain (Baldwin et al., 1993; Baldwin et al., 1994). Additional questions about temporal stability, contextual variation, factors influencing activation, and so on, need to be studied more extensively to develop a complete theory of working models.

We acknowledge, however, that the alternative explanations regarding continuity and unreliability cannot be ruled out completely. Research is needed to examine who tends to rate self according to which attachment description and under what circumstances. In the meantime, researchers should be wary about administering the categorical attachment measure at a different time from their other dependent measures.

References

Ainsworth, M. D. S., Blehar, M. S., Waters, E., & Wall, S. (1978). *Patterns of attachment: A psychological study of the strange situation.* Hillsdale, NJ: Erlbaum.

Baldwin, M. W. (1992). Relational schemas and the processing of social information. *Psychological Bulletin, 112,* 461–484.

Baldwin, M. W., Carrell, S. E., & Lopez, D. F. (1990). Priming relationship schemas: My advisor and the Pope are watching me from the back of my mind. *Journal of Experimental Social Psychology, 26,* 435–454.

Baldwin, M., Fehr, B., Keedian, E., Seidel, M., & Thomson, D. (1993). An exploration of the relational schemas underlying attachment styles: Self-report and lexical decision approaches. *Personality and Social Psychology Bulletin, 19,* 746–754.

Baldwin, M. W., & Holmes, J. G. (1987). Salient private audiences and awareness of the self. *Journal of Personality and Social Psychology, 52,* 1087–1098.

Baldwin, M. W., Keelan, J. P. R., Fehr, B., & Koh-Rangarajoo, E. (1994). Social cognitive conceptualization of attachment styles: Availability and accessibility effects. Manuscript submitted for publication.

Barnes, M. (1991). "I like your style": Effects of visualization and memory on relationship attachment styles. Unpublished honors thesis, University of Winnipeg.

Bartholomew, K. (1993, August). *Assessment of stability in adult attachment representations.* Paper presented at the meeting of the American Psychological Association, Toronto.

Bartholomew, K. (1994). Assessment of individual differences in adult attachment. *Psychological Inquiry, 5,* 23–67.

Bartholomew, K., & Horowitz, L. M. (1991). Attachment styles among young adults: A test of a four-category model. *Journal of Personality and Social Psychology, 61,* 226–244.

Blatt, S. J., & Homann, E. (1992). Parent–child interaction in the etiology of dependent and self-critical depression. *Clinical Psychology Review, 12,* 47–91.

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

Bowlby, J. (1973). *Attachment and loss: Vol 2. Separation, anxiety, and anger.* New York: Basic.

Bowlby, J. (1980). *Attachment and loss: Vol 3. Loss, sadness, and depression.* New York: Basic.

Brennan, K. A., & Morris, K. A. (1993, August). *Attachment and personality disorders among children of divorce.* Presentation at the annual meeting of the American Psychological Association, Toronto.

Brennan, K. A., & Shaver, P. R. (1990). Dimensions of adult attachment and the dynamics of romantic relationships. Unpublished manuscript, Department of Psychology, State University of New York at Buffalo.

Brennan, K. A., Shaver, P. R., & Tobey, A. E. (1991). Attachment styles, gender and parental problem

drinking. *Journal of Social and Personal Relationships, 8,* 451–466.

Bretherton, I. (1985). Attachment theory: Retrospect and prospect. *Monographs of the Society for Research in Child Development, 50,* 3–35.

Bretherton, I. (1990). Communication patterns, internal working models, and the intergenerational transmission of attachment relationships. *Infant Mental Health Journal, 11,* 237–252.

Cantor, N., & Kihlstrom, J. F. (1987). *Personality and social intelligence.* Englewood Cliffs, NJ: Prentice-Hall.

Cicchetti, D. V., & Sparrow, S. A. (1981). Developing criteria for establishing interrater reliability of specific items: Applications to assessment of adaptive behavior. *American Journal of Mental Deficiency, 86,* 127–137.

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

Collins, N. L., & Read, S. J. (1990). Adult attachment, working models, and relationship quality in dating couples. *Journal of Personality and Social Psychology, 58,* 644–663.

Crittenden, P. M. (1990). Internal representational models of attachment relationships. *Infant Mental Health Journal, 11,* 259–277.

Eiden, R. D., Leonard, K., & Senchak, M. (1993, August). *Prediction of marital adjustment from couples' attachment styles.* Presentation at the annual meeting of the American Psychological Association, Toronto.

Feeney, J. A., & Noller, P. (1990). Attachment style as a predictor of adult romantic relationships. *Journal of Personality and Social Psychology, 58,* 281–291.

Feeney, J. A., & Noller, P. (1991). Attachment style and verbal descriptions of romantic partners. *Journal of Social and Personal Relationships, 8,* 187–215.

Feeney, J. A., Noller, P., & Callan, V. J. (1994). Attachment style, comunication and satisfaction in the early years of marriage. In K. Bartholomew & D. Perlman (Eds.), *Advances in personal relationships* (Vol. 5, pp. 269–308). London: Jessica Kingsley.

Feeney, J. A., Noller, P., & Hanrahan, M. (1994). Assessing adult attachment: Developments in the conceptualization of security and insecurity. In M. B. Sperling & W. H. Berman (Eds.), *Adult attachment: Conceptual and clinical applications across the life span* (pp. 128–152) New York: Guilford.

Griffin, D. W., & Bartholomew, K. (1994). The metaphysics of measurement: The case of adult attachment. In K. Bartholomew & D. Perlman (Eds.), *Advances in personal relationships* (Vol. 5, pp. 17–52). London: Jessica Kingsley.

Hammond, J. R., & Fletcher, G. O. (1991). Attachment styles and relationship satisfaction in the development of close relationships. *New Zealand Journal of Psychology, 20,* 56–62.

Hazan, C., Hutt, M. J., & Markus, H. (1991). Continuity and change in inner working models of attachment. Unpublished manuscript, Department of Human Development, Cornell University.

Hazan, C., & Shaver, P. R. (1987). Romantic love conceptualized as an attachment process. *Journal of Personality and Social Psychology, 52,* 511–524.

Hazan, C., & Shaver, P. R. (1990). Love and work: An attachment-theoretical perspective. *Journal of Personality and Social Psychology, 59,* 270–280.

Hazan, C., & Shaver, P. R. (1994a). Attachment as an organizational framework for research on close relationships. *Psychological Inquiry, 5,* 1–22.

Hazan, C., & Shaver, P. R. (1994b). Deeper into attachment theory. *Psychological Inquiry, 5,* 68–79.

Hendrick, C., & Hendrick, S. S. (1994). Attachment theory and close relationships. *Psychological Inquiry, 5,* 38–41.

Higgins, E. T., Bond, R. N., Klein, R., & Strauman, T. (1986). Self-discrepancies and emotional vulnerability: How magnitude, accessibility and type of discrepancy influence affect. *Journal of Personality and Social Psychology, 51,* 5–15.

Horowitz, M. (1979). *States of mind.* New York: Plenum.

Keelan, J. P. R., Dion, K. L., & Dion, K. K. (1994). Attachment style and heterosexual relationships among young adults: A short-term panel study. *Journal of Social and Personal Relationships,* 11, 201–214.

Kirkpatrick, L. A., & Davis, K. E. (1994). Attachment style, gender, and relationship stability: A longitudinal analysis. *Journal of Personality and Social Psychology, 66,* 502–512.

Kirkpatrick, L. A., & Hazan, C. (1994). Attachment styles and close relationships: A four-year prospective study. *Personal Relationships, 1,* 123–142.

Kirkpatrick, L. A., & Shaver, P. R. (1992). An attachment-theoretical approach to romantic love and religious belief. *Personality and Social Psychology Bulletin, 18,* 266–275.

Kobak, R. (1994). Adult attachment: A personality or relationship construct? *Psychological Inquiry, 5,* 42–44.

Kobak, R. R., & Hazan, C. (1991). Attachment in marriage: Effects of security and accuracy of working models. *Journal of Personality and Social Psychology, 60,* 861–869.

Lamb, M. E., Thompson, R. A., Gardner, W. P., Charnov, E. L., & Estes, D. (1984). Security of infantile attachment as assessed in the "strange situation": Its study and biological interpretation. *Behavioral and Brain Sciences, 7,* 127–171.

Levy, M. B., & Davis, K. E. (1988). Lovestyles and attachment styles compared: Their relations to each other and to various relationship characteristics. *Journal of Social and Personal Relationships, 5,* 439–471.

Lewis, M. (1994). Does attachment imply a relationship or multiple relationships? *Psychological Inquiry, 5,* 47–51.

Main, M., & Goldwyn, R. (1988). Adult attachment classification system. Unpublished manuscript, University of California, Berkeley.

Main, M., Kaplan, N., & Cassidy, J. (1985). Security in infancy, childhood, and adulthood: A move to the level of representation. *Monographs of the Society of Research in Child Development, 50,* 67–104.

Mikulincer, M., Florian, V., & Tolmacz, R. (1990). Attachment styles and fear of personal death: A case study of affect regulation. *Journal of Personality and Social Psychology, 58,* 273–280.

Mikulincer, M., & Nachshon, O. (1991). Attachment styles and patterns of self-disclosure. *Journal of Personality and Social Psychology, 61,* 321–331.

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

Noller, P., & Feeney, J. A. (1994). Whither attachment theory: Attachment to our caregivers or to our models? *Psychological Inquiry, 5,* 51–56.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

Pietromonaco, P. R., & Carnelley, K. B. (1994). Gender and working models of attachment: Consequences for perceptions of self and romantic relationships. *Personal Relationships, 1,* 63–82.

Pistole, M. C. (1989). Attachment in adult romantic relationships: Style of conflict resolution and relationship satisfaction. *Journal of Social and Personal Relationships, 6,* 505–510.

Russell, J. A. (1989). Affect grid: A single-item scale of pleasure and arousal. *Journal of Personality and Social Psychology, 57,* 493–502.

Scharfe, E., & Bartholomew, K. (1994). Reliability and stability of adult attachment patterns. *Personal Relationships, 1,* 23–43.

Senchak, M., & Leonard, K. E. (1992a). Attachment styles and marital adjustment among newlywed couples. *Journal of Social and Personal Relationships, 9,* 51–64.

Senchak, M., & Leonard, K. E. (1992b). Stability of spouses' attachment styles over the first year of marriage. Unpublished raw data.

Shaver, P. R., & Brennan, K. A. (1992). Attachment styles and the "Big Five" personality traits: Their connections with each other and with romantic relationship outcomes. *Personality and Social Psychology Bulletin, 18,* 536–545.

Shaver, P. R., Collins, N., & Clark, C. L. (in press). Attachment styles and internal working models of self and relationship partners. In G. J. O. Fletcher & J. Fitness (Eds.), *Knowledge structures in close relationships: A social psychological approach.* Hillsdale, NJ: Erlbaum.

Shaver, P. R., & Hazan, C. (1993). Adult romantic attachment: Theory and evidence. In W. H. Jones & D. Perlman (Eds.), *Advances in personal relationships* (Vol. 4, pp. 29–70). London: Jessica Kingsley.

Simpson, J. A. (1990). Influence of attachment styles on romantic relationships. *Journal of Personality and Social Psychology, 59,* 971–980.

Sprecher, S., Aron, A., Hatfield, E., Cortese, A., Potapova, E., & Levitskaya, A. (1994). Love: American style, Russian style, and Japanese style. *Personal Relationships, 1,* 349–369.

Sroufe, L. A., & Fleeson, J. (1985). Attachment and the construction of relationships. In W. W. Hartup & Z. Rubin, (Eds.), *Relationships and development* (pp. 51–71). Hillsdale, NJ: Erlbaum.

Tidwell, M., Cooper, M. L., & Shaver, P. (1993, August). *Attachment and stressful life events as predictors of distress.* Presentation at the annual meeting of the American Psychological Association, Toronto.

Tinio, E. J. (1992). Career and relationship commitment in two university samples: The importance of self-efficacy. Unpublished honors thesis, University of Winnipeg.

West, M., Sheldon, A., & Reiffer, L. (1987) An approach to the delineation of adult attachment: Scale development and reliability. *The Journal of Nervous and Mental Disease, 175,* 738–741.