THE METAPHYSICS OF MEASUREMENT: 
THE CASE OF ADULT ATTACHMENT

Dale W. Griffin and Kim Bartholomew

According to the Alfie Singer theory of romance, relationships, like sharks, must constantly move forward or die. Of course, this preference for novelty and change is an exaggerated, one-sided view of romantic relationships. Happy romances also have periods of calm and stability, when partners integrate their past experiences and their new discoveries into a deeper understanding of each other and their partnership. It seems to us that the same pattern can be found in psychological research on close relationships: periods of exhilarating discovery, when a novel theoretical approach leads a variety of researchers to attack a particular problem, followed by a period of reassessment and integration, in which the independent empirical discoveries are examined in light of a broader understanding.
Research on adult romantic attachment patterns has so far moved relentlessly forward in the manner of the shark. New correlates of adult attachment patterns seem to be reported each month. Following Hazan and Shaver’s (1987) demonstration that Ainsworth’s classification system (based on infants’ relationships with their caregivers) could be extended to adults, the study of ‘attachment style’ became perhaps the fastest growing sub-field in the area of close relationships. Attachment style has been shown to be related to (among other things) jealousy, paternal drinking, sexual activity, relationship satisfaction, conflict styles, coping responses, neuroticism, and depression (see Shaver & Hazan, 1993, for a comprehensive review).

However, even though this ‘age of discovery’ has yielded a wealth of empirical results, it has lacked two important ingredients. First, the basic theoretical logic of attachment theory as developed by John Bowlby and Mary Ainsworth has become more and more remote from the strongly empirical nature of the research, and second, there has been no common methodology for defining and measuring attachment style – new assessment instruments have generally been derived empirically, usually through factor-analytic methods. Hazan and Shaver (in press; Shaver & Hazan, 1993) have recently called for a renewed focus on the theoretical tenets of attachment theory. They urge researchers interested in adult attachment to take the theory, not the measures, as their starting point (Hazan & Shaver, in press). Our position is similar to that of Hazan and Shaver; however, we go one step further. For although we agree that research should be guided by theory, we stress that a theory can never be tested directly, only indirectly through observations on specific measuring instruments. Any test of any theory is limited by the quality of the measurement procedures used in that test. Furthermore, the choice of a measurement procedure carries with it implicit theoretical assumptions about the nature of the phenomenon under study. Thus, our position, and the theme of this chapter, is that choosing an approach to measuring adult attachment is a central theoretical decision, and determines not only how well a theory will be tested but what theory is actually tested.

THE ROLE OF MEASUREMENT

Whenever we teach a course on research methods, the first class inevitably includes the following four points. First, psychology is a science because of the methods it uses, not because of the subject matter it includes. Second, the methods of science are based on an apparently simple principle: assigning numbers to events in a way that is public and replicable. Third, when psychologists assign numbers to events, they rarely are able to directly measure the phenomenon of interest (e.g., ‘love’) but instead measure observed variables (e.g., pencil marks on a 7-point love scale) that correspond, somewhat imperfectly, to the true, underlying concept. Finally, we argue, research based on even the most exciting, novel, and penetrating theory is useless – worse, misleading – unless there is close correspondence between measured variables and the underlying theoretical constructs. We also tell our students that a rigorous consideration of measurement issues is often bypassed by researchers because it seems dry and boring next to the simple pleasure of theoretical musings or the tingling excitement of data collection. In our classes, we do our best to persuade the researchers of tomorrow that measuring with measurement issues is not only a sine qua non of good social science but can also be intellectually rewarding. We hope to get the same message across in this chapter.

Perhaps another line from Alfy Singer will help to bring this point to life. Singer, the comedian protagonist of Woody Allen’s Annie Hall, tells an audience about the time he was suspended from college for cheating on a Metaphysics exam – he was caught looking into the soul of the person sitting next to him. It is the absurdity of the contrast between marks on a piece of paper (the source of ordinary exam cheating) and the soul of a human being (the source of this cheating scandal) that makes the joke funny. Yet that same absurd distance between pencil marks on a piece of paper and feelings such as love, happiness, hope, and depression must be bridged by techniques of psychological measurement. Theories may be about love and happiness, but until psychologists are able to look directly into people’s minds – or their souls – data must be collected at the level of pencil marks.

OVERVIEW

In any question of individual differences, there are choices to be made among qualitatively different approaches to measurement, as well as among specific measuring instruments. In this chapter we will not be comparing specific measuring instruments, nor will we be comparing specific theories of adult attachment. Instead we will focus on broader questions about the nature of individual differences in adult attachment. First, is the underlying construct best characterized as a set of types or categories, or as continuous dimensions? If a categorical approach is chosen, are people to be placed into discrete groups or fuzzy sets? In the field of adult attachment, researchers have operationalized ‘attachment patterns’ in all three ways. Adult attachment has been measured by (and thus implicitly theorized to be) two and three continuous dimensions (e.g., Collins & Read, 1990; Simpson, Rholes & Nelligan, 1992), three or four discrete categories (e.g., Hazan & Shaver, 1987; Main, Kaplan & Cassidy,
1985), and continuous ratings of four prototypical attachment patterns (Bartholomew & Horowitz, 1991). In the pages that follow, we will explore the theoretical and empirical implications of these distinctions, and suggest some methods of conceptual and statistical analysis to answer the questions they raise. We will use a variety of data sets that have been collected to test Bartholomew's (1990) four-category model to illustrate our points.

We begin by defining and discussing the three prominent approaches to conceptualizing and measuring individual differences in adult attachment: dimensional, typological, and prototype. Next, we outline a particular theory of adult attachment, Bartholomew's four-category model, that incorporates all three approaches, and we describe the associated measurement instruments. Following this, we define and illustrate some conceptual questions that arise within and between the measurement approaches: How can the 'correct' attachment dimensions be identified? Are these dimensions reducible to general personality factors? Are prototype measures of attachment more powerful than dimensional measures? Do attachment categories have emergent properties? Are prototype measures of attachment more powerful than categorical measures? And what implications do the different approaches hold for the assessment of reliability and stability? Throughout each section, we emphasize the interconnection of theoretical analysis and measurement procedure.

CONCEPTUALIZING ADULT ATTACHMENT: DIMENSIONS, TYPES, AND PROTOTYPES

Dimensional Approaches

Psychologists most commonly measure individual differences by multi-item self-report scales. Typically, these are bipolar scales and they characterize an individual as a point along a continuous dimension. Most often, a combination of dimensions is used. An individual's thoughts, feelings, or behavior are explained by a linear combination of her or his standing on the dimensional measures. A dimensional approach implies, first, that people can only be quantitatively ordered: There is no sudden qualitative shift at any point on the dimension that divides people into different categories. Second, it implies that each dimension has an effect independent of the other dimensions, that is, there is no interaction between the various dimensions. Dimensional measures are not as common as categorical approaches in adult attachment research, but it has been proposed that two fundamental dimensions underlie adult attachment patterns (Griffin & Bartholomew, 1993; Shaver & Hazan, 1993). In addition, continuous measures have been developed to assess various components of adult attach-
of interest. Furthermore, this implies that only between-group variance is meaningful and within-group variance is essentially random error. A rare example of a strong grouping theory that makes these assumptions explicit is Gangestad and Snyder’s (1985) argument for the discrete nature of self-monitoring, based on a genetic theory about the origins of personality traits they argue that – on the self-monitoring dimension at least – there really are only two ‘types’ of people in the world, high self-monitors and low self-monitors. According to Gangestad and Snyder, each discrete type has a qualitatively different strategy for managing social interaction.

Most researchers use grouping approaches because of their convenience (‘types as labels’), not because of the measurement implications (‘types as distinct forms’; see Block, 1961). And indeed, referring to a relatively few types makes communication very economical, whether the groups are Type A’s versus Type B’s or securely versus insecurely attached individuals. The advantages of economical communication are matched by advantages in statistical analysis. Measuring discrete groups allows the researcher to analyze mean differences by way of Analysis of Variance, and to summarize the data effectively using tables and figures of group profiles. The most profound advantage of a grouping approach is, of course, that it might capture the ‘true’ nature of the phenomenon in question. For example, Hazan and Shaver (in press) propose that different attachment groups use qualitatively different strategies of emotional regulation. The strategy associated with each group should translate into a different profile of variables. In this way, a few groups or categories can serve as a shorthand summary of a complex pattern of individual differences – providing, of course, that all members of a group share the same pattern or profile (see, e.g., Weinberger & Schwartz, 1990).

The disadvantages of a grouping approach spring directly from a researcher’s failure to test or discuss its implicit assumptions. It is a fundamentally different view of human nature to postulate different kinds of people than it is to think of people varying along underlying dimensions. If a dimensional approach is more appropriate theoretically, then the advantages of economy must be weighed against the disadvantages of losing valuable within-group variance. One compromise approach is to measure finer sub-group differences within the major groups (Ainsworth, Blehar, Waters & Wall, 1978), though in practice these sub-groups are often too small to be useful.

The simplicity of group-based statistical analysis can have unfortunate consequences. Comparing group means via Analysis of Variance yields a temptation to think about the grouping variable – the so-called ‘independent variable’ – in causal terms; a temptation more easily resisted (or at least more easily spotted) in the correlational analyses characteristic of a dimensional approach.

Another, more invidious problem with grouping approaches is the natural human tendency to use stereotypes to simplify information-processing. Thirty years of social psychological research on the cognitive bases of stereotypes has demonstrated that once people are classified into groups, observers tend to exaggerate the similarity between members of a category, minimize the differences between members, and overlook evidence that disconfirms the simple classification scheme. These cognitive and perceptual biases are just as likely to affect the way that scientists or clinicians think about members of a diagnostic category as the way that lay people think about members of distinctive minority groups (Horowitz, Wright, Lowenstein & Parad, 1981). Thus, a grouping scheme may not only lead to overly simplistic and generalized thoughts about group members, but even to stereotyped courses of therapy that treat individuals as members of a diagnostic category rather than as unique individuals. A classification scheme that explicitly recognizes (and measures) variations within a category is the prototype approach.

Prototype Approaches

According to prototype theory (Rosch, 1978), most natural categories such as ‘birds’ do not have distinct boundaries as implied by classical all-or-none grouping approaches to categorization. Instead, birds vary from the most prototypical or central category members (robin) to the least prototypical (ostrich). A prototype is an ideal category member defined in terms of the most common features of members of that category, with no particular features being individually necessary or jointly sufficient to define group membership. Group members therefore differ in the degree to which they correspond to the group exemplar or prototypical group member. From this perspective, categories are thus ‘fuzzy sets’ that may overlap with one another and that contain members with varying degrees of typicality.

The prototype approach to categorization allows for the complex patterns of individual differences that may define ‘types’ of persons, while also recognizing that not all members of a group are equally good exemplars of that group. This approach thereby integrates and addresses the limitations of both dimensional and grouping approaches to the measurement of individual differences, and has been used in classifying personality types, psychological symptoms, and diagnostic categories (e.g., Cantor, Smith, French & Mezzich, 1980; Horowitz et al. 1981). Prototype-based measurement seems particularly appropriate for attachment research. It is unlikely that many adults correspond perfectly to any one attachment pattern, given the multitude of past influences (e.g., genetic predispositions and life experiences) and present situational and relationship-specific influences on
their orientations to attachment relationships. Rather, over time and across situations, most adults would be expected to show varying degrees of two or more attachment patterns. A prototype approach to attachment allows for the assessment of how well an individual fits each prototype at any one time and how much that individual’s fit varies over time.

We should note that there are dimensional approaches that implicitly employ the concept of prototypes. In particular, Block’s method of ‘criterion’ Q sorts, in which individual Q sorts are compared to an idealized Q sort generated to represent a particular construct, is conceptually similar to defining a prototypical group member against which particular individuals are compared. The Q sort methodology has been used productively in the attachment field. Waters and Deane have developed an attachment Q sort for young children that yields a score indicating the correspondence of a given child’s behavior with experts’ judgments of the behavior of a prototypical secure child (Waters & Deane, 1985). Similarly, Kobak has developed a Q sort for assessing the attachment strategies of adults along two dimensions (Kobak & Hazan, 1991).

Bartholomew (1990; Bartholomew & Horowitz, 1991) has developed a four-category model of adult attachment that explicitly uses a prototype approach. As we show below, this model can be expressed in terms of all three forms of measurement: dimensional, grouping, and prototype. We therefore use this model to organize our empirical demonstrations.

THE FOUR-CATEGORY MODEL

Bowlby (1973) proposed that individuals internalize their experiences with caretakers such that they form internal working models of their own self-worth and of their expectations of care and support from others. Bartholomew has systematized Bowlby’s conception of internal working models by defining individual differences in adult attachment in terms of the intersection of two dimensions – positivity of the self model and positivity of models of hypothetical others. Dichotomizing each dimension as positive or negative then leads to four prototypical attachment patterns (see Figure 1). See Bartholomew (1990), Bartholomew and Horowitz (1991), and Bartholomew (1993) for more detailed descriptions of the model. Previous research has confirmed that the four attachment prototypes can be reliably measured, that a two-dimensional structure underlies the four patterns as hypothesized, and that different methods of assessment converge as expected (Bartholomew & Horowitz, 1991; Griffin & Bartholomew, 1993).

Although each of the three approaches to defining individual differences is implicit in the four-category model, the model is explicitly based on the prototype approach to measurement and categorization. Each of the four attachment patterns identified by the model is conceptualized as a theoretical ideal or prototype with which individuals may correspond to varying degrees. This approach also recognizes that most individuals exhibit elements of more than one attachment pattern and that to assess adequately individuals' feelings, expectations, and behaviors in the attachment domain, it is necessary to consider their profiles across the four attachment patterns. For example, an individual who is moderately secure but secondarily dismissing would look quite distinct from an individual who is moderately secure and secondarily preoccupied. When using semi-structured interviews as the basis of assessment (as in Bartholomew & Horowitz, 1991), raters are trained to understand the four prototypes and to assess subjects' degree of correspondence with each of the four prototypes on 9-point scales. Similarly, subjects may rate their own degree of similarity with written descriptions of the four attachment prototypes.

Although the four-category model is theoretically based on the intersection of two underlying dimensions (positivity of self models and of other models), Bartholomew’s assessment procedure does not include a direct measure of the two underlying dimensions. Rather, ratings of the two

---

**Figure 1: Four-category model of adult attachment**

- **Model of Self (Anxiety)**
  - Positive (Low)
  - Negative (High)
- **Model of Other (Avoidance)**
  - Secure
  - Dismissing
  - Fearful
  - Preoccupied

---

The Metaphysics of Measurement

Model of Self (Anxiety)

Positive (Low)  
Negative (High)

Model of Other (Avoidance)

Secure  
Dismissing  
Fearful  
Preoccupied
dimensions can be derived by linear combinations of the four prototype ratings. For instance, the positivity of the self model is derived by adding together ratings of the patterns defined by positive self models (the secure and dismissing) and subtracting the ratings of the patterns defined by negative self models (the fearful and preoccupied). Based on the resulting values for the self and other dimensions, individuals can be placed within the space defined by the intersection of the self- and other-model dimensions; the four attachment ratings can thus be reduced to a single point in a two-dimensional space.

The purest measure of the underlying dimensions would be obtained by having raters (or subjects themselves) code the two dimensions directly. In fact, Bartholomew originally attempted to develop coding protocols to measure the two dimensions, but ran into two related problems. First, it was surprisingly difficult to develop simple rules for defining being high or low on the two dimensions because the two dimensions are not independent in their effects, but interactive. In other words, there are different ways of showing a positive self model depending on the valence of the other model and different ways of showing a positive other model depending on the valence of the self model. In addition, even trained raters found it very difficult to rate subjects on these two underlying dimensions; in contrast, they found it relatively much easier to rate subjects in terms of their fit with the four prototypes. Interestingly, Mayer and Bower (1986) have demonstrated that untrained lay persons are remarkably adept at learning complex personality prototypes.

Within the four-category model, the highest of the four attachment ratings can also be used to classify subjects into an attachment category. The resulting group membership thus indicates only the best-fitting attachment pattern; the specific rating on that attachment pattern would indicate how well that category describes a given individual. In addition to assessing the degree of typicality of group members, this method of assigning group membership avoids the problem of deciding on appropriate cut-offs when dimensional measures are used to define category membership. Although individuals can be categorized within the four-category model, in practice few individuals are ideal exemplars of a particular attachment pattern; rather, most display elements of two or more attachment patterns. For example, of a group of 77 subjects (Study 1, Bartholomew & Horowitz, 1991), the average rating of subjects in their best-fitting attachment group was approximately 6 on a 9-point scale of degree of fit with the corresponding attachment prototype. Moreover, only 52 per cent of the subjects were at least two points higher on their highest attachment rating (and, therefore, best-fitting category) than on the next best-fitting attachment prototype and only 14 per cent were given a score of 7 or higher indicating that they were good to excellent exemplars of their best-fitting category.

Four distinct measuring instruments have been developed in the context of the four-category model. Three of these instruments directly assess the match between an individual and each of the attachment prototypes. The Peer Attachment Interview is a semi-structured one-hour interview in which participants are asked to describe their friendships, romantic relationships, and feelings about the importance of close relationships. On the basis of interview audio recordings, expert judges then rate (on a 9-point scale) how closely each participant corresponds to each of the four attachment prototypes. Interjudge reliabilities are typically greater than .90. A parallel Family Attachment Interview follows a similar format and rating procedure, but focuses on representations of experiences in the family of origin rather than on peer relationships. The Relationship Questionnaire (RQ) consists of four short paragraphs describing the attachment prototypes as they apply to close peer relationships (see Appendix A). Participants are asked to rate (on a 7-point scale) how well they correspond to each prototype. (See Bartholomew & Horowitz, 1991, for more information about these procedures.)

A third instrument, the Relationship Scales Questionnaire (RSQ), is an indirect measure of the prototypes. The RSQ consists of 30 phrases drawn from the paragraph descriptions in Hazan and Shaver’s (1987) attachment measure, Bartholomew and Horowitz’s (1991) Relationship Questionnaire, and Collins and Read’s (1990) Adult Attachment Scale (see Appendix B). Participants rate (on a 5-point scale) how well each item fits their characteristic style in close relationships. RSQ scores for the four attachment prototypes are derived by computing the mean of the items representing each prototype. Four items contribute to the score for the preoccupied and fearful pattern whereas five items contribute to the scores for the other two patterns.

The internal consistencies of the RSQ prototype scores are variable and at times quite low, in one sample ranging from alpha = .41 for the secure pattern to alpha = .70 for the dismissing pattern. This follows neither from the low number of items making up each prototype score, nor from some accidental psychometric flaw in the construction of the scales. Instead, the internal consistencies of the RSQ scales can be low because two orthogonal dimensions (self-model and other-model) are being combined. Consider, for example, two of the items making up the secure subscale: ‘I am comfortable depending on other people’ reflects the other-model attachment dimension whereas ‘I worry about being alone’ (reverse scored) reflects the self-model dimension.

Interestingly, despite their factorial complexity, the RSQ pattern scores still show convergent validity. Table 1 presents correlations (N=153) between interview prototype ratings and 1) the direct self-report prototype ratings of the RQ, and 2) the indirect self-report pattern ratings of the RSQ.
Table 1. Correlations Between Interview Attachment Ratings and RQ and RSQ Self-Report Ratings

<table>
<thead>
<tr>
<th>Interview Ratings of Adult Attachment Prototypes</th>
<th>RQ Ratings</th>
<th>RSQ Subscales</th>
</tr>
</thead>
<tbody>
<tr>
<td>Secure</td>
<td>.22</td>
<td>.25</td>
</tr>
<tr>
<td>Fearful</td>
<td>-.33</td>
<td>-.31</td>
</tr>
<tr>
<td>Preoccupied</td>
<td>-.16</td>
<td>-.15</td>
</tr>
<tr>
<td>Dismissing</td>
<td>-.09</td>
<td>-.08</td>
</tr>
<tr>
<td>Fearful</td>
<td>-.24</td>
<td>-.40</td>
</tr>
<tr>
<td>Secure</td>
<td>.13</td>
<td>.15</td>
</tr>
<tr>
<td>Fearful</td>
<td>.50</td>
<td>.32</td>
</tr>
<tr>
<td>Preoccupied</td>
<td>.02</td>
<td>.16</td>
</tr>
<tr>
<td>Dismissing</td>
<td>.27</td>
<td>.02</td>
</tr>
<tr>
<td>Preoccupied</td>
<td>.33</td>
<td>.34</td>
</tr>
<tr>
<td>Dismissing</td>
<td>.40</td>
<td>.26</td>
</tr>
<tr>
<td>Dismissing</td>
<td>.17</td>
<td>.06</td>
</tr>
<tr>
<td>Secure</td>
<td>.00</td>
<td>.14</td>
</tr>
<tr>
<td>Fearful</td>
<td>.26</td>
<td>.44</td>
</tr>
<tr>
<td>Preoccupied</td>
<td>.47</td>
<td></td>
</tr>
<tr>
<td>Dismissing</td>
<td>.10</td>
<td></td>
</tr>
</tbody>
</table>

It is noteworthy that for both the RQ and RSQ the convergent validity coefficients for the secure pattern was the lowest of the four patterns. This finding is suggestive that security of attachment may be especially susceptible to self-report biases. Interestingly, the ratings of the dismissing pattern, a defensive style that some have suggested may be inherently difficult to assess with self-report measures (e.g., Hazan & Shaver, 1987), showed very respectable convergence across methods. However, overall the modest magnitude of the convergent correlations indicates that the interview and self-report methods of measurement are far from identical.

We also examined correlations at the dimensional level. For the interview ratings and self-report scales, self-model dimension scores were derived by adding together subjects’ scores on the patterns hypothesized to represent positive self models (the Secure and Dismissing patterns) and subtracting subjects’ scores on the patterns hypothesized to represent negative self models (the Preoccupied and Fearful patterns). Similarly, other-model dimension scores were derived by adding together subjects’ scores on the patterns hypothesized to represent positive other models (the Secure and Preoccupied patterns) and subtracting subjects’ scores on the patterns hypothesized to represent negative other models (the Dismissing and Fearful patterns). See Griffin and Bartholomew (1993) for further details on this technique.

When we move to the dimensional level of analysis, the convergent correlations are somewhat higher. The interview measure of the self-model dimension correlated .41 with the RQ self-model and .37 with the RSQ self-model, and the interview measure of the other-model dimension correlated .46 with the RQ other-model and .48 with the RSQ other-model. The mean absolute value of the correlations for the non-corresponding dimensional measures was only .10.

**APPROACHES TO ASSESSING ADULT ATTACHMENT: SIX MEASUREMENT ISSUES**

1. Defining the ‘Right’ Attachment Dimensions

The most common procedure for constructing dimensional measures is a two-step process: first, a large pool of relevant items are generated (usually based on an equal measure of theoretical guidance and convenient borrowing), and second, the ‘underlying structure’ of the items is deduced by some method of dimensional reduction (usually factor analysis). This procedure has the distinct advantage of producing a few highly reliable scales from a larger set of items. Furthermore, when principal components’ extraction with orthogonal rotation is used (as is typical), the resulting scales have low intercorrelations, simplifying subsequent data analysis. However, the factor-analytic approach also carries with it the temptation to regard the derived factors as somehow representing the ‘fundamental structure’ of the domain of interest, when in fact they represent the researcher’s choice of items to be included in the analysis. It is easy to confuse the process of uncovering the structure of a given set of items and uncovering the structure of the human psyche, and this confusion is multiplied by the tempta-
tion to believe that by naming a factor one has captured its true meaning (the so-called 'nomological fallacy').

Consider the following example in the attachment domain. One common method of generating a pool of attachment items is to take the descriptions of the different attachment groups defined by Hazan and Shaver (1987) and use each phrase in the descriptions as a different item. When these items are analyzed by a principal components analysis with either orthogonal or oblique rotation, a clear two-dimensional solution is found (Simpson et al. 1992; see also Feeney, Noller & Callan, this volume).

One dimension or factor consists of items relating to the Avoidance of Intimacy and the other consists of items relating to Anxiety. This pattern of results is highly reproducible; when we gathered data from 650 undergraduate students and analyzed (using principal components analysis for extraction, and an oblique rotation to simple structure) the same 13 items as Simpson et al. (1992) used, we found exactly the same factor structure. Eleven of the thirteen items had loadings of at least .5 on their hypothesized factor and none had loadings of greater than .20 on their non-corresponding factor. The first factor was anchored positively by Avoidance items such as 'I am nervous when anyone gets too close' and negatively by 'I find it easy to get close to others'. The Anxiety factor contained items such as 'I worry that romantic partners won't want to stay with me'. These two factors correspond closely to the self-model and other-model dimensions said by Bowlby to underlie attachment patterns (Griffin & Bartholomew, 1993).

It might seem from this evidence that adult attachment can be well described by two dimensions of anxiety (or positivity of the self model) and avoidance (or positivity of the other model). But there is another level at which the attachment descriptions can be factor-analyzed – the prototype descriptions themselves can be used as items and their intercorrelations entered into a data-reduction program. Brennan, Shaver, and Tobey (1991) did just this when they asked students to rate themselves on Hazan and Shaver's three attachment patterns and Bartholomew and Horowitz's (1991) four attachment patterns, and submitted the resulting data matrix to a principal components analysis. The obtained two-dimensional solution was very clear: the first factor was anchored positively by ratings of Bartholomew's and Hazan's secure styles and negatively by ratings of Bartholomew's fearful style and Hazan's avoidant style; the second factor was anchored positively by Bartholomew's preoccupied and Hazan's anxious-ambivalent styles and negatively by Bartholomew's dismissing style.

This pattern, too, is easily replicable. We asked our sample of 650 undergraduate students to fill out Bartholomew and Horowitz's four-category Relationship Questionnaire (RQ) as well as the multi-item Relationship Scales Questionnaire (the RSQ). We then entered the four RQ prototype scores with the four derived RSQ prototype scores into a principal compo-

The Metaphysics of Measurement

ments analysis. As expected, two clear factors resulted: one running from secure to fearful ratings, and the other from preoccupied to dismissing ratings (represented by the dotted lines in Figure 1). Conceptually, these dimensions represent a 45 degree rotation from the self-model and other-model (or anxiety and avoidance) dimensions found with the individual items.

Which of these dimensional structures is 'correct', 'natural', or 'fundamental'? This is a theoretical, not an empirical question, and therefore it cannot be answered by factor analysis alone. Data reduction techniques serve only to simplify the set of items that were selected by the researcher. It is this selection process that determines whether the majority of the items line up with the self- and other-model dimensions (as in the item level of the RSQ), or with secure – fearful or preoccupied – dismissing dimensions (as in the prototype ratings). A dimension is only 'correct' in regard to a particular theoretical position: it takes theoretically-based demonstrations of construct validity (Cronbach & Meehl, 1955) to demonstrate that the dimensions obtained are 'right' in regard to that particular theory. See Wiggins (1991) for an elegant example of validating a 'fundamental' dimensional structure on the basis of cultural, philosophical, psychological, and linguistic evidence.

In a recent paper, we demonstrated the construct validity of the self- and other-model dimensions underlying the four-category model (Griffin & Bartholomew, 1993). We demonstrated that the two attachment dimensions showed convergent validity across methods: that is, self-reports, peer-reports, partner reports, and expert raters’ judgments of the dimensions intercorrelated highly. We also demonstrated that the different methods of rating the attachment dimensions showed discriminant validity: relatively low correlations were found between the two different dimensions rated by the same method. Most important, we demonstrated that the two attachment dimensions were related to theoretically relevant outcome variables. In this case, based on Bowlby’s conception of the attachment dimensions, we hypothesized that the self-model dimension would be highly related to the positivity of one’s self-concept whereas the other-model dimension would be highly related to the positivity of one’s interpersonal orientation.

Figure 2 displays the results of the latent variable analysis used to test this hypothesis. The rectangles in the figure represent measured variables, whereas the large circles represent the hypothetical latent variables ('constructs') underlying the measured variables. The small circles represent the error terms for the measured variables, thus the double-headed arrows represent the 'method correlations' between measured variables using the same method of measurement. A structural modeling approach using latent variables has a number of advantages in theory testing. Unlike exploratory
factor analysis or principal components analysis, structural modeling with latent variables is a confirmatory technique that requires theory-guided decisions about measurement issues (such as which variables load on which factors) before the analysis begins. Furthermore, programs based on maximum-likelihood estimation of structural models (such as LISREL, EQS, or EZPATH) provide significance tests to evaluate how well – more accurately, how badly – the hypothesized model fits the observed data. Another advantage to these techniques is that they allow the simultaneous consideration of the measurement model (i.e., how well the different methods measure the underlying construct) and the structural model (the relations among the latent, ‘true’ variables). This enables the researcher to estimate the relation between the conceptual variables underlying the measured variables, controlling for random measurement error and systematic method variance. For further discussion of structural equation modelling see Connell and Tanaka (1987).

The consideration of Figure 2 reveals at least three major conclusions. First, all three methods of measuring the attachment dimensions (self-report, interview ratings, and peer ratings) do a reasonable job, as indicated by the moderate to high loadings on the two latent attachment dimensions. Nevertheless, the loadings of the individual variables on the dimensions are far enough from 1.0 to attenuate dramatically the observed correlations between measured variables. Second, there is significant method variance ('correlated errors') between self-report attachment measures and a self-report measure of interpersonal orientation. A similar relation exists for peer reports. Thus, attempts to assess the relations between the attachment dimensions and the outcome dimensions by correlating individual measures are confounded both by random error and by systematic method variance. Third, at the level of latent constructs, the attachment self-model dimension is all but perfectly related to the self-concept variable, and the attachment other-model dimension is similarly highly related to the interpersonal orientation variable.

Thus, the attachment dimensions are highly related to outcome variables that are specified by theory to match each dimension. In this sense, we can say that we are using the 'right' attachment dimensions. But it is still important to demonstrate that the construct of adult attachment adds explanatory power above and beyond standard individual difference measures. In the next two sections, we demonstrate that the self- and other-model dimensions are not simply reducible to fundamental dimensions of personality, and that the four attachment prototypes are not simply reducible to the two attachment dimensions.
2. The Discriminant Validity of the Attachment Dimensions

Perhaps the biggest measurement problem in the field of close relationships is the common failure to consider the discriminant validity of new measures. This practice has led to the proliferation of largely redundant measures distinguished by different names. Such a practice makes it very difficult to assess and integrate the findings of different research groups. In the field of personality research, integration has proceeded more quickly than in the field of close relationships. As a result, personality research offers a well-validated "frame of reference" against which to evaluate and compare new constructs; the five-factor model of personality, commonly referred to as the 'Big Five' (e.g., Digman, 1990; John, 1990; McCrae & Costa, 1987; Wiggins & Pincus, 1992). The Big Five personality traits are five dimensions that commonly appear in factor analyses of both natural language trait terms and personality scale items. The five-factor model provides a widely recognized taxonomy of personality dimensions – Neuroticism (N), Extroversion (E), Openness to Experience (O), Agreeableness (A), and Conscientiousness (C) – that appear to underlie most personality constructs (Piedmont, McCrae & Costa, 1991, p.630). Furthermore, 'Correlation of a scale with measures of the five basic factors of personality provides an efficient way to understand its psychological meaning and orient it in a widely shared conceptual system' (Piedmont et al., 1991, p.636). Thus, it is important to examine 1) how the attachment dimensions correlate with the traits identified by the five-factor model, 2) if the attachment dimensions can be explained by a combination of the Big Five dimensions, and 3) whether the attachment dimensions can explain interpersonal functioning above and beyond the Big Five.

Table 2 presents correlations between the five basic personality scales, as indexed by the short form of the NEO Personality Inventory (Costa & McCrae, 1985), and two different self-report measures of the attachment dimensions collected from a sample of 470 introductory psychology students. Inspection of Table 2 reveals that the dimensional measures derived from the RQ and RSQ showed very similar profiles of correlations with the trait measures. Both measures of the self-model dimension were highly correlated with the Neuroticism Scale of the NEO and modestly correlated with Extroversion, Openness, and Agreeableness. Both measures of the other-model dimension were moderately correlated with Extroversion, and had slightly lower correlations with Openness. In all cases, the RSQ measure had higher correlations with the relevant Big Five scale than did the RQ measure, despite the low internal consistency of the RSQ pattern measures.

<table>
<thead>
<tr>
<th>NEO Subscales</th>
<th>RQ Dimensions</th>
<th>RSQ Dimensions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Self model</td>
<td>Other model</td>
</tr>
<tr>
<td>Neuroticism</td>
<td>-.49</td>
<td>-.13</td>
</tr>
<tr>
<td>Extroversion</td>
<td>.25</td>
<td>.30</td>
</tr>
<tr>
<td>Openness</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>Agreeableness</td>
<td>.25</td>
<td>.24</td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>.18</td>
<td>.01</td>
</tr>
</tbody>
</table>

Note: N = 470; correlations greater than |.15| are significant at p<.01.

The RQ and RSQ dimensional measures showed moderate convergent validity (convergent correlations were .58 for the self-model dimension and .57 for the other-model dimension). This suggests that there is considerable error variance in these self-report measures – enough to attenuate markedly the observed correlations between the attachment dimensions and the NEO scales. Thus, we examined a structural model in which the five personality scales predicted two latent outcome variables: a self-model and other-model latent variables each defined by an RQ and an RSQ measured variable. This enabled us to assess what proportion of the reliable or ‘true’ variance in the attachment measures was explained by the Big Five dimensions. Figure 3 presents the results of this analysis, using the EQS program of structural equation modeling (Steiger, 1989). The personality scales explained part, but not nearly all, of the reliable variance in the attachment dimensions: 48 per cent of the variance in the latent self-model variable was explained by the Big Five and 27 per cent of the variance in the latent other-model variable was explained by the Big Five. Clearly, there is much more to the attachment dimensions than what can be explained by the fundamental dimensions of personality.

An examination of the individual path coefficients in Figure 3 reveals that by far the single largest predictor of the self-model dimension was Neuroticism, with a standardized path coefficient of -.60. This is consistent with the notion that a positive self-model is associated with a lack of anxiety and self-doubt. A typical item in the Neuroticism scale is 'Sometimes I feel completely worthless’. The single largest predictor of the other-model dimension was Extroversion but this relation was not particularly strong (standardized path coefficient = .39). In fact, none of the Big Five scales captures the element of comfort with intimacy that is the marker of a positive model of others. Although some of the items on the Extroversion
scale measure sociability (e.g., 'I really enjoy talking to people'), many focus on a sense of energy and surgency (e.g., 'I am a very active person' and 'I often feel as if I'm bursting with energy'), which should not relate highly to the other model dimension. Agreeableness, also moderately related to the other-model dimension, is defined by items such as 'I try to be courteous to everyone I know', which also fail to capture comfort versus avoidance of intimacy.

We have demonstrated that the attachment dimensions are not reducible to the fundamental factors identified in personality measures, but this does not speak to whether the attachment dimensions are useful above and beyond the five factors. We expect that attachment measures would uniquely predict behavior and emotion in intimate relationships. Ideally, then, we should examine whether the attachment dimensions, over and above the Big Five factors, predict such variables as individuals' reports of relationship quality, feelings of satisfaction in their relationship, or the number and success of past relationships. However, such variables are not included in this data set. Probably the most relevant measure available is a short form of the Interpersonal Dependency Inventory (Hirschfeld, Klerman, Gough, Barrett, Korchin & Chodoff, 1977). One subscale, Emotional Reliance on others, involves both a negative self-concept and a sense of trust and thus is negatively related to the attachment self-model dimension of the RQ, \( r(468) = - .50, p < .001 \), and positively related to the attachment other-model dimension of the RQ, \( r(468) = .23, p < .001 \). A second dependency subscale, Assertion of Autonomy, involves a defensive rejection of the need for others, and thus is positively related to the attachment self-model dimension, \( r(468) = .33, p < .001 \), and negatively related to the attachment other-model dimension, \( r(468) = -.33, p < .001 \). A third subscale, Lack of Social Self-Confidence, measures an individual's feelings of uncertainty and insecurity in social relations, and thus is negatively related to the attachment self-model dimension, \( r(468) = -.26, p < .001 \), although it is unrelated to the attachment other-model dimension, \( r(468) = .04, ns \).

Although the dependency subscales are correlated with the measures of the attachment dimensions in predictable ways, do these relationships remain even after the Big Five scales are accounted for? To answer this question, we first regressed each dependency subscale onto the Big Five scales and then examined whether the two attachment dimensions from the RQ significantly added to the prediction equation. The Emotional Reliance subscale was well-predicted by the Big Five scales \( (R^2 = .55) \), but even so, the two attachment dimensions added significantly to the prediction equation \( (AR^2 = .15, p < .001) \). Furthermore, both the self-model and other-model dimensions had significant partial regression coefficients. We found the same pattern for the Assertion of Autonomy subscale: it was very strongly predicted by the Big Five scales \( (R^2 = .52) \), the two attachment dimensions still added significantly \( (AR^2 = .13, p < .001) \), and both self-model and other-model dimensions remained significant in their own right. The attachment dimensions were not, however, uniquely related to the Lack of Social Self-Confidence subscale. After the Big Five scales were accounted for \( (R^2 = .67) \), the attachment dimensions added only 1 per cent to the variance explained. Overall, then, for two of the three dependency scales, we found evidence for the 'incremental validity' of the attachment dimensions in predicting different types of dependency. Not only is there considerable variance in the attachment dimensions that cannot be accounted for by the Big Five, but this additional variance is meaningfully related to a relevant interpersonal variable, dependence on others. We should note that the same general conclusions were drawn by Shaver and Brennan (1992) in their recent study relating the Big Five scales to their 3-category model of attachment.
With this evidence in hand—evidence of the construct validity and incremental utility of the dimensions underlying the four-category model of adult attachment—we are now in a position to ask: is that all there is? Are the prototype measures really necessary or are the underlying dimensions sufficient to describe adult attachment?

3. The Explanatory Power of Attachment Prototypes

According to the theoretical predictions of the four-category model, the two attachment dimensions alone are not enough to describe an individual's attachment orientation. Instead, knowing how well people correspond to all four attachment prototypes should help explain their behavior in intimate relationships over and above the predictable value of the attachment dimensions. Bartholomew and Horowitz (1991) established the predictive validity of the attachment prototypes by demonstrating that interview ratings of each attachment prototype were related to distinct patterns of interpersonal problems (as assessed by both self and peer reports). These specific patterns did not merely reflect a particular point on the self- or other-model dimension, but represented a behavioral strategy that resulted from the particular combination of underlying dimensions. When we reanalyzed Study 1 of Bartholomew and Horowitz (1991) in terms of the attachment dimensions, a different (and, we think, oversimplified) picture emerged than suggested by the prototype analyses. Consider, for example, the tendency to be cold and distant with others, as measured by the Cold subscale of the Inventory of Interpersonal Problems (IIP; Horowitz, Rosenberg, Barer, Ureno & Villasenor, 1988). Self-reported problems with coldness were unrelated to the self-model dimension and highly negatively related (r = -.60) to the other-model dimension. These dimensional results gave no hint that the problems with coldness were primarily associated with how well an individual fit the dismissing prototype (r = .58), secondarily associated with a poor fit to the secure (r = -.27) and preoccupied prototypes (r = -.35), and not at all related to the individual's rating on the fearful prototype (r = .03).

Clearly, then, the measurement of attachment prototypes adds "interpretational power" beyond the measurement of attachment dimensions. However, does it add to predictive power? To answer this question, we examined the relation between self-reported attachment variables and self-reported relationship satisfaction in 77 couples who were assessed as part of a longitudinal study (Bartholomew & Scharle, 1993). For women, it appeared that the attachment prototypes did not add to the predictive power of the attachment dimensions. When the two derived attachment dimensions from the RQ were entered into a regression equation predicting relationship satisfaction, the dimensions together accounted for 14 percent of the variance (p<.001) whereas the four prototype ratings added only 1 percent. For men, the prototype ratings were much more useful than the dimension scores: the two derived attachment dimensions accounted for only 2 percent of the variance in relationship satisfaction, whereas the four prototype ratings increased the R-square by 28 percent.

The source of these discrepant results can be pinpointed by examining the correlations between the individual prototype ratings and ratings of relationship satisfaction. For women, the pattern of correlations is quite simple: a good fit to the secure prototype is positively related to satisfaction and a fit to any of the insecure prototypes is negatively related to satisfaction. This fits a dimensional structure because a high rating on the one prototype that represents both a positive self model and a positive other model has an opposite effect from high ratings on any prototype that represents at least one negative model. For men, too, only ratings of the secure prototype were positively correlated with satisfaction whereas ratings of all three insecure prototypes were negatively correlated with satisfaction. But where for women the negative correlations between all three insecure prototypes and satisfaction were of comparable magnitude, for men the insecure prototypes differed in importance: the fit to the dismissing prototype was very important (r = -.52, p<.01), followed by the fit to the preoccupied (r = -.28, p<.05) and the fearful (r = -.11, ns) This pattern of correlations does not match the ordering expected if the dimensions were sufficient to predict relationship satisfaction. Instead it indicates the special status of the dismissing prototype in understanding male self-reports of relationship satisfaction. In sum, whether or not a given variable is significantly or even highly related to the attachment dimensions, the prototype measures may add both predictive power and interpretational clarity.

These findings hint that each of the four discrete conjunctions (high/high, high/low, low/high, and low/low) of self and other models have configurural or emergent properties not captured by linear combinations of the underlying dimensions. Does this mean that different 'types' of people can be identified by their best-fitting attachment prototype? This question, the existence of typologies, brings us to the grouping or categorical approach to measuring adult attachment.

4. The Nature of Attachment Categories: The Search for Emergent Properties

The four-category model can be used to assign individuals to a single best-fitting prototype. We have argued that such a categorization oversimplifies the measurement of attachment and leads to the loss of important information about how well the individual fits all four types. We will pursue this argument further in the next section below. But it is still important to
The Metaphysics of Measurement

tations of being let down from coming true. A similar pattern was found for problems with passivity, but here the effect is primarily that of the fearful group being uniquely high, and the other three groups being statistically indistinguishable by post-hoc tests.

![Graphs showing correlations between different groups and problem-solving]

Figure 4. Scores on IIP subscales by attachment group
Note: S = Secure; P = Preoccupied; F = Fearful; D = Dismissing.
Interpersonal problem scores are standardized.

Correlational results generally mirror the patterns found in the group comparisons, although the sub-samples are too small to allow significance tests. Correlations (calculated after controlling for sex) between problems with expressiveness and scores on the other-model dimension derived from interview ratings were stronger for individuals with low self models (r =
.56) than for individuals with high self models \((r = .21)\). This captures the fact that the two groups with low self models (preoccupied and fearful) have very different strategies with regard to expressiveness, whereas the two groups with high self models do not differ so much in this domain. Similar evidence of the moderating effect of level of self model was found for problems with being exploited: the correlation with the other-model dimension was stronger for individuals with low self models \((r = -.26)\) than for individuals with high self models \((r = .03)\); and for problems with passivity: for individuals with negative self model, \(r = -.33\) whereas for individuals with a positive self model, \(r = -.10\). Overall, then, there is at least suggestive evidence for the configurational properties of the four attachment types.

A typological approach may have a place in the study of adult attachment, but at what cost? Or, as Mendelsohn et al. (1982) forcefully argue, 'typological analyses are committed to assumptions of discontinuity that are probably false and thus to analytic procedures that both obscure information and throw information away. Throwing away information is often a good idea, but one wants to know what it is before doing so'. In the next section, we examine the importance of the information that is 'thrown away' by a categorical approach.

### 5. Categories Versus Prototypes: The Utility of Within-Group Ratings

To illustrate the loss of information entailed by a categorical approach, we once again turned to our large sample \((N=470)\) of self-report questionnaires. We chose as our dependent variables two outcome measures that should be simply and strongly related to the underlying attachment dimensions: anxiety, as measured by the Trait Anxiety Inventory (Spielberger, 1980), and interpersonal problems with warmth, as measured by the IIP. We first demonstrate that self-reports ratings of the four attachment prototypes on the RQ add predictive power above that given by group membership. Of course, it is not surprising that four continuous scales are more predictive than a discrete measure of group membership, given the greater reliability of the rating scales. More interesting and important is the second step, where we assess the magnitude and meaningfulness of the within-group correlations between attachment dimensions and the outcome variables.

Group membership accounted for 18 percent of the variance in reported anxiety, and ratings of the four attachment prototypes added another 9 percent \((p<.001)\) to the predictability of this variable. Similarly, group membership explained 17 percent of the variance in interpersonal problems with warmth, and ratings of the four attachment prototypes again added 9 percent \((p<.001)\). The within-group correlations between the two outcome variables and the two attachment dimensions derived from the RQ are presented in Table 3. An examination of the table reveals consistent evidence of the utility of within-category variation on the attachment prototypes. In fact, many of these 'attenuated' correlations still manage to break the .3 barrier. For example, consider those 15 per cent of the students who rated the dismissing prototype as best representing their attachment pattern. Even within this highly selected group, the within-group variance on the two attachment dimensions was quite strongly related to the outcome variables.

<table>
<thead>
<tr>
<th>Attachment Category</th>
<th>Anxiety</th>
<th>Problems with Warmth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Secure</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self Model</td>
<td>-.27**</td>
<td>.06</td>
</tr>
<tr>
<td>Other Model</td>
<td>.06</td>
<td>.25**</td>
</tr>
<tr>
<td>Fearful</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self Model</td>
<td>-.26**</td>
<td>-.14</td>
</tr>
<tr>
<td>Other Model</td>
<td>.06</td>
<td>.27**</td>
</tr>
<tr>
<td>Preoccupied</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self Model</td>
<td>-.48**</td>
<td>.10</td>
</tr>
<tr>
<td>Other Model</td>
<td>-.32*</td>
<td>.41**</td>
</tr>
<tr>
<td>Dismissing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self Model</td>
<td>-.33**</td>
<td>.29**</td>
</tr>
<tr>
<td>Other Model</td>
<td>-.20</td>
<td>.40**</td>
</tr>
</tbody>
</table>

Note: \(N\)'s within categories range from 64 to 152.

\(* p < .05; ** p < .01\).

This finding has implications for both theory and practice. Theoretically, it underlines our argument that two individuals with the same primary attachment strategy may have quite different interpersonal experiences if they have different secondary or even tertiary strategies. In practice, it means that even when the researcher is faced with a special subpopulation that does not include the whole range of attachment groups, that researcher can still expect the four prototype ratings to yield meaningful results. For example, in a sample of men in treatment for wife assault, there were few...
6. Assessing 'True Scores': The Role of Reliability and Stability

One apparently simple way to determine the quality of a measuring instrument is to evaluate whether the scores it provides are 'true' or reliable. However, such an evaluation is not as straightforward as it seems. Reliability refers only to the proportion of true variance (relative to error variance) in a measuring instrument— it gives no guarantee that the 'true scores' being measured correspond to the particular theoretical construct of interest. Thus, reliability assessment does not speak to the meaning of observed scores, only to their repeatability. For dimensional approaches to attachment, repeatability is inferred indirectly from the 'internal consistency' of the multi-item scale. Internal consistency (usually indexed by coefficient alpha, Cronbach, 1951) refers to the degree of intercorrelation among the different items making up a scale, and can be assessed based on a single administration (e.g., see Feeney et al. this volume). The reliability of continuous single-item measures, whether dimensional or prototype, can be assessed in one of two ways. Most commonly, researchers administer the same self-report attachment form twice a few weeks apart, and then correlate these scores (e.g., Levy & Davis, 1988). Less commonly, researchers have multiple judges rate subjects' attachment patterns based on the same information (such as an attachment interview, e.g., see Bartholomew & Horowitz, 1991). The latter strategy is conceptually equivalent to correlating scores on alternative forms of the same questionnaire at approximately the same time (Nunnally, 1978). The reliability of categorical measures is also assessed through test-retest or alternate-form reliabilities, but in this case is indexed by the concordance of group membership across measures or times (e.g., see Steele & Steele, this volume). Note that percent agreement is not sufficient to index the reliability of categorical measures; a coefficient of concordance such as kappa, that takes the base rate occurrence of each category into account, is necessary (Cohen, 1968).

No matter how reliability is assessed, a multi-item attachment measure generally is more reliable than a comparable single-item measure. This is true whether the aggregation occurs over multiple items on a self-report measure or over multiple raters of an interview. This is an important advantage of using multi-item measures and multiple judges. Problems with unreliability found in continuous single-item measures are exacerbated in the case of single-item categorical measures because the 'either-or' format is highly sensitive to even small changes in response tendency.

For single-item attachment measures, it is important to distinguish between the information given by a test-retest correlation and that given by an alternate-form correlation. Correlation or concordance (in the case of categorical measures) across time is a function of both the reliability of the measure (that is, the proportion of 'true score' in that measure) and the stability of the underlying construct (that is, the tendency for people to change differentially over time). Thus, a low test-retest correlation may indicate low reliability of the measure, low stability of the construct, or both. In general, high reliability is necessary for high stability, but even reliably measured phenomena can be unstable. For example, positive or negative mood may be reliably measured at any one time but would not necessarily remain stable over time.

There are at least three general strategies for determining the role of reliability and stability in a given cross-time correlation or concordance. First, the researcher may have evidence about both the test-retest correlation and the reliability of the measure as assessed by internal consistency, parallel forms, or multiple raters. The stability of the construct can then be understood in light of the reliability of the measured variable. For example, Feeney et al. (this volume) interpret a test-retest correlation of a dimensional measure of attachment in light of the internal consistency of the measure. Similarly, in his review of studies showing intergenerational transmission of attachment patterns using categorical measures, Van IJzendoorn (1992) points out that the concordance among raters at a given time (assessing reliability only) helps us to interpret the meaning of the concordance in attachment patterns between parents and their children (assessing reliability plus intergenerational consistency).

A second general strategy is to separate formally the reliability of the measured variables and the stability of the underlying construct through the use of structural equation modeling. This requires multiple measures of the construct at each time point and is generally appropriate only for continuously measured variables, either dimensions or prototypes. For example, Scharfe and Bartholomew (in press) examined the stability of Bartholomew's two dimensions by using multiple raters of the Peer Attachment Interview at each time point and found high stability (.72 to .85) over an eight-month time period. A third strategy is to compare the magnitude of the test-retest correlation or concordance over different intervals of time. For example, Scharfe and Bartholomew (1993) observed that test-retest correlations for dimensional measures of attachment were comparable over periods of two weeks to nine months. Such a finding implies an unreliable measurement of a fairly stable underlying construct.
CONCLUSION: NAMING DOES NOT MAKE IT SO

Poets and philosophers have long struggled with the problem of identifying essential properties. Juliet mused 'What's in a name? A rose by any other name would smell as sweet.' A more recent poet proclaimed 'If it looks like a duck, and walks like a duck, and quacks like a duck, it's a duck.' And the philosopher and cognitive scientist Douglas Hofstadter wryly declared 'Shakespeare's plays weren't written by him, but by someone else with the same name.' Underneath these flowery phrases and wry words lies a deep problem. In empirical psychology, we rarely stop our data-gathering to wrestle with abstract notions such as essential qualities. Yet every time we design a questionnaire or a symptom checklist, we should give at least some thought to the metaphysics of measurement. For example, what lies beneath the name 'attachment style'? What is its true form or essential qualities? Is it a set of categories that defines different types of people, or a combination of continuous dimensions along which people vary, or a set of prototypic ideals to which people can be compared? Depending on the theory one starts with, attachment style could be defined in any of these ways. And each of these approaches carries with it specific assumptions. In this chapter, we have illustrated how to turn these assumptions into specific testable questions.

Overall, the answers to the questions we asked here support the prototype approach to measuring adult attachment (Bartholomew, 1990). That is, although we showed that the dimensions of self-model and other-model underlying Bartholomew's model were highly related to relevant outcome variables even after the five fundamental dimensions of personality were accounted for, we found that prototype measures still added predictive and interpretational power to the dimensions. Furthermore, although we showed that the four categories of attachment specified by Bartholomew's model seemed to have 'emergent properties' not predictable from the underlying dimensions, we found that a person's degree of fit to a given category or prototype was also vitally important in understanding her or his interpersonal functioning. At this point, however, we think it is too early to pick a 'winner' among the different approaches to measuring adult attachment. The area of research is very young, and is best served by a variety of approaches, each with its own particular advantages and costs. But we hope that this chapter will encourage researchers in this area to rely on their theoretical positions to define the measurement procedures they use, rather than vice versa.

In this chapter, we have focused on three different approaches to conceptualizing individual differences in adult attachment patterns. However, the same kinds of arguments could be, and need to be, made in terms of different methods of measurement. For instance, theory may tell us the degree to which individuals are expected to be consciously aware of their attachment models and, therefore, whether self-report or other forms of measurement are more appropriate. Similarly, our theoretical definition of attachment models should dictate the degree of specificity (i.e., a single relationship) or generality of our measures of attachment (see Collins & Read, this volume). Methods of measurement, like approaches to distinguishing attachment styles, should be chosen on the basis of theory, rather than ease. Furthermore, these theoretically based choices need to be explicitly acknowledged and their validity systematically tested. After all, theories are never tested directly, and any empirical test can never be better than the quality of the measurement used. Theories may be the heart and soul of psychology, but measures are the flesh and bones. Progress in the field of adult attachment will only be assured if we attend to the needs of the flesh as well as to those of the soul.

NOTES

1. The 153 individuals referred to are actually members of 77 heterosexual couples (one subject deleted because of missing data) assessed by Bartholomew and Scharfe (1993). Although it is inappropriate to infer the statistical significance of dependent (i.e., couple-level) correlations on the basis of individual-level data (Kenny, 1988), we have presented the individual correlations for simplicity. The pattern of results is very similar when the correlations are computed separately by sex.

2. This study was conducted to test one particular model – not to rule out or invalidate other conceptualizations. To our knowledge a formal test of the construct validity of other attachment dimensions has not been carried out.

REFERENCES


The Metaphysics of Measurement


APPENDIX A

Relationship Questionnaire (RQ) Items

Secure pattern
It is easy for me to become emotionally close to others. I am comfortable depending on them and having them depend on me. I don’t worry about being alone or having others not accept me.

Fearful pattern
I am uncomfortable getting close to others. I want emotionally close relationships, but I find it difficult to trust others completely, or to depend on them. I worry that I will be hurt if I allow myself to become too close to others.

Preoccupied pattern
I want to be completely emotionally intimate with others, but I often find that others are reluctant to get as close as I would like. I am uncomfortable being without close relationships, but I sometimes worry that others don’t value me as much as I value them.

Dismissing pattern
I am comfortable without close emotional relationships. It is very important to me to feel independent and self-sufficient, and I prefer not to depend on others or have others depend on me.

Notes: Each paragraph is rated on a 7-point scale ranging from 'Not at all like me' to 'Very much like me'. The measure can be worded either in terms of general orientations to close relationships, orientations to romantic relationships, or orientation to a specific relationship (with others' changed to a specific partner ‘P’). The measure can also be reworded in the third person and used to rate others' attachment patterns (e.g., close friends or romantic partners).
APPENDIX B

Relationship Scales Questionnaire (RSQ) Items

1. I find it difficult to depend on other people. (Fear)
2. It is very important to me to feel independent. (Dis)
3. I find it easy to get emotionally close to others. (Sec)
4. I want to merge completely with another person.
5. I worry that I will be hurt if I allow myself to become too close to others. (Fear)
6. I am comfortable without close emotional relationships. (Dis, Pre-R)
7. I am not sure that I can always depend on others to be there when I need them.
8. I want to be completely emotionally intimate with others. (Pre)
9. I worry about being alone. (Sec-R)
10. I am comfortable depending on other people. (Sec)
11. I often worry that romantic partners don’t really love me.
12. I find it difficult to trust others completely. (Fear)
13. I worry about others getting too close to me.
15. I am comfortable having other people depend on me. (Sec)
16. I worry that others don’t value me as much as I value them. (Pre)
17. People are never there when you need them.
18. My desire to merge completely sometimes scares people away.
19. It is very important to me to feel self-sufficient. (Dis)
20. I am nervous when anyone gets too close to me.
21. I often worry that romantic partners won’t want to stay with me.
22. I prefer not to have other people depend on me. (Dis)
23. I worry about being abandoned.
24. I am uncomfortable being close to others. (Fear)
25. I find that others are reluctant to get as close as I would like. (Pre)
26. I prefer not to depend on others. (Dis)
27. I know that others will be there when I need them.
28. I worry about having others not accept me. (Sec-R)
29. Romantic partners often want me to be closer than I feel comfortable being.
30. I find it relatively easy to get close to others.

Notes: Items scored on a 5-point scale ranging from ‘not at all like me’ to ‘very like me’. Items that make up subscales for the attachment patterns defined by the four-category model are marked as follows: Sec=Secure, Fear=Fearful, Pre=Preoccupied, and Dis=Dismissing. R indicates reversed scoring.