This is a preprint of an article published in [Green, Donald Philip, Susan L. Goldman, and Peter Salovey. 1993. Measurement Error Masks Bipolarity in Affect Ratings. *Journal of Personality and Social Psychology* 64:1029-41. Reprinted 2003 in G.J. Boyle and D.H. Saklofske (eds.) *Individual Differences*. London: Sage Publications]. Pagination of this preprint may differ slightly from the published version.

Measurement Error Masks Bipolarity in Affect Ratings

Donald Philip Green, Susan Lee Goldman, and Peter Salovey

For years, affect researchers have debated about the true dimensionality of mood. Some have argued that positive and negative moods are largely independent and can be experienced simultaneously. Others claim that mood is bipolar, that joy and sorrow represent opposite ends of a single dimension. The 3 studies presented in this article suggest that the evidence that purportedly shows the independence of seemingly opposite mood states, that is, low correlations between positive and negative moods, may be the result of failures to consider biases due to random and nonrandom response error. When these sources of error are taken into account using multiple methods of mood assessment, a largely bipolar structure for affect emerges. The data herein speak to the importance of a multimethod approach to the measurement of mood.

For some time now, it has been popular for investigators of mood and emotion to assert that positive and negative affect are independent. The idea is that someone can experience joy tinged with sorrow, hatred tempered by love, and anger coincident with kindness. A foray into the recent mood literature reveals conclusions such as the following: “Periodic factor analyses. . . have produced a strongly similar pattern of results over the years: two large factors, one for positive and the other for negative affect” (Moore & Isen, 1990, pp. 4-5). The view that positive and negative affect are not opposite ends of a single dimension (i.e., are not strongly negatively correlated) but, instead, represent nearly orthogonal dimensions of mood can be traced to several highly influential studies of well-being by Bradburn and his colleagues (Bradburn, 1969; Bradburn & Cap-lovitz, 1965).

We suspect that the empirical assumptions on which the “independent factors” view of positive and negative mood rest are unsound. In our view, the independence of positive and negative affect is a statistical artifact. The conclusion that positive and negative affect are largely uncorrelated fails to take account of the errors of measurement that arise in mood assessment—errors that occur because people have difficulty translating their moods into survey responses. We argue that widely used methods of mood assessment are statistically unreliable and that, moreover, the measurement error associated with the assessment of positive and negative feelings is not random. Instead, the errors of measurement in both scales tend to be correlated because they emerge from the same sources. We shall demonstrate that when random and nonrandom measurement error is taken into account, the independence of positive and negative affect, however defined, proves ephemeral. When one adjusts for random and systematic error in positive and negative affect, correlations between the two that at first seem close to 0 are revealed to be closer to -1.00 and support a largely bipolar structure.

Donald Philip Green, Department of Political Science, Yale University; Susan Lee Goldman and Peter Salovey, Department of Psychology, Yale University.

This article benefited considerably from comments by Edward Diener, Jay Hull, John D. Mayer, Alexander Rothman, David Watson, and four anonymous reviewers regarding a draft. Thanks to Jon Cowden, who assisted in the preparation of the manuscript.

This research was supported by a grant to Donald Philip Green from Yale’s Institution for Social and Policy Studies, by a National Science Foundation Minority Predoctoral Fellowship to Susan Lee Goldman, and by the following grants to Peter Salovey: National Institutes of Health Grants BRSG S07 RR07015 and CA 42101 and National Science Foundation Grant BNS-9058020.

The survey instruments used to collect the mood data used in this article as well as the data themselves are available from us on request.

Correspondence concerning this article should be addressed to Donald Philip Green, Department of Political Science, Yale University, P.O. Box 3532, Yale Station, New Haven, Connecticut 06520-3532, or to Peter Salovey, Department of Psychology, Yale University, P.O. Box 11A, Yale Station, New Haven, Connecticut 06520-7447.

Measurement of Mood

Historically, early affect researchers assumed the intuitively appealing idea that positive and negative moods represent opposite poles of one underlying dimension. To these investigators, it seemed likely that a happy person was one who was not sad and that a sad individual could not simultaneously be happy. Such notions led Guilford (1954), echoing Wundt (1897), to declare that “an affective scale is a bipolar one” (p. 264). Affect was thought to be essentially the same as the first dimension of the semantic differential—evaluation—with similar pleasant and unpleasant poles (Osgood, Suci, & Tannenbaum, 1957).

However, early factor analytic work rarely confirmed this bipolar formulation. When investigators asked subjects to rate mood adjectives, pleasant and unpleasant items tended not to load on opposite ends of one dimension but rather formed two separate dimensions, each with a single pole. It seemed that people either felt pleasant or not or felt unpleasant or not, and their ratings of positive and negative affect seemed largely independent (e.g., Borgatta, 1961; Clyde, 1963; McNair & Lorr, 1964; Nowlis & Nowlis, 1956). These investigators assumed that bipolarity could only be found when it was forced onto the data by deliberately labeling the ends of rating scales with adjectives that were antonyms, but that when affect was measured with only a single adjective for each scale item, separate positive and negative dimensions would be recovered (R. F. Green & Goldfried, 1965). Consequently, various mood scales were developed on the basis of the view that affect was best characterized by at least two monopolar factors (e.g., Izard's, 1972, Differential Emotions Scale; McNair, Lorr, & Droppleman's, 1971, Profile of Mood States, and Nowlis's, 1965, Mood Adjective Checklist).

It was not long, however, before the independent factors view was challenged. Bentler (1969) constructed an adjective version of the semantic differential containing 141 evaluative terms. Two hundred subjects were asked to describe the emotional meaning of 200 different concepts using these terms (as well as other terms loading on the other two traditional dimensions of the semantic differential, activity and potency; there were 352 total adjectives). This massive matrix was subjected to nonmetric multidimensional scaling (Bentler, 1966) to identify the latent ordinal dimensions associated with observed adjective ratings. Once the biasing effects of response style were offset by controlling for the number of adjectives checked by a subject, a pattern of correlations emerged that supported a bipolar structure. Positively valenced adjectives were inversely correlated with negatively valenced ones.

In the years that followed, investigators identified other systematic sources of error variance that, when accounted for, transformed data that appeared to contain independent, monopolar positive and negative affect factors into dimensions that were in fact better characterized by bipolarity. For instance, Meddis (1972) identified a set of systematic sources of error variance that tends to operate in adjective rating scales, namely that (a) many contained a "don't know" category that was not really in the middle of the scale (e.g., Thayer, 1967), which subjects often used to indicate noncomprehension of the adjective or its irrelevance to the rating task and (b) many adjective rating scales were not symmetrical (e.g., McNair & Lorr, 1964), often containing two or three levels of acceptance (e.g., *extremely, quite a bit, and a little)* and only one level of rejection (e.g., *not at all).* Meddis (1972) claimed that such asymmetries in scale construction were "suppressing negative correlations and prejudicing the factor analysis against the discovery of bipolar factors" (p. 180). After correcting for such sources of systematic error, Meddis (1972) obtained strong evidence for a small set of bipolar mood factors.

The pièce de résistance for the view that affect is bipolar was a now classic study whose title declared, simply, "Affect Space is Bipolar" (Russell, 1979; see also related work, Russell, 1978; Russell & Mehrabian, 1977). Russell (1979) argued that bipolarity was suppressed in most studies of affect because of a series of measurement issues that created systematic error, including those described by Meddis (1972), and (a) the sample of emotion words included on scales often underrepresented one end of a bipolar continuum, (b) instructions often asked subjects to rate how they feel over extended periods of time, (c) response formats often resulted in bimodal rather than normal distributions for each item (the modal response was often *not at all),* and (d) items in close proximity on the scale often showed spuriously inflated intercorrelations. After correcting these shortcomings, Russell (1979) found strong evidence that these defects in measurement had previously obscured the fact that affect space is bipolar and defined, in part, by a strong, primary bipolar factor, Pleasure-Displeasure, and a secondary bipolar factor, Aroused-Sleepy.

After 20 years, during which the original independent view was largely replaced by this bipolar perspective, it was somewhat unexpected that the 1980s would be characterized by a resurgence of interest in a model of affect claiming relative independence of positive and negative factors. Three independent laboratories published prominent articles in fairly quick succession, all providing evidence that, at least under certain conditions, mood was best characterized by two broad independent dimensions, positive affect and negative affect (Diener & Emmons, 1984; Warr, Barter & Brownbridge, 1983; Zevon & Tellegen, 1982).

Zevon and Tellegen (1982), taking an ideographic approach to the study of mood, asked a small group of subjects to complete a daily 60-item mood adjective checklist for 90 consecutive days. P-factor analytic techniques were used to analyze the data, and each subject's first two factors were derived and compared (using within-individual or P-factor analysis). In 21 of their 23 cases, these factors were largely independent positive and negative affect dimensions. (These factors were not in a strict sense monopolar; their opposite ends reflected the adjectives *sleepy* and *calm)* Zevon and Tellegen (1982) characterized these two factors as "descriptively bipolar but affectively unipolar dimensions" (p. 112). These investigators did not see their results as necessarily contradicting those offered by the "bipolar" camp. Rather, they viewed the issue as, in part, a rotational one. If one chose to rotate the Zevon and Tellegen (1982) positive and negative dimensions 45°, a bipolar Pleasant-Unpleasant dimension (and a unipolar arousal dimension) would result. (See Watson, 1988, pp. 128-129, for a discussion of how dimensional labeling varies with factor rotation.) And as Watson (1988) argued in several studies building on the Zevon and Tellegen (1982) approach, the selection of adjectives determines the factor analytic solution and hence the emergence of independent or bipolar mood factors (see also Larsen & Diener, 1992; Watson & Tellegen, 1985).

Another perspective promoting the "independence" view was offered by Diener and Emmons (1984). They noted that positive and negative affect could be strongly and inversely related at any given moment in time and yet still be independent in terms of how people reflect on their moods over a longer period of time. Using daily mood reports that extended from 30 to 70 days, they discovered that positive and negative affect were inversely correlated only during intense emotional experiences. However, as a longer and longer time frame was considered (from moments, to days, to 3-week intervals), the correlation between positive and negative affect decreased dramatically. Diener and Emmons (1984) concluded that under conditions of low intensity and over longer time frames, positive and negative affect might not be polar opposites (see also Diener & Iran-Nejad, 1986, for evidence supporting this view and Watson, 1988, for evidence against it), and mood might be best characterized by dimensions of intensity of affect and frequency of positive and negative affect (Diener, Larsen, Levine, & Emmons, 1985).

A third perspective on the independence of positive and negative affect was provided by Warr et al. (1983). They argued that because good and bad events do not tend to be associated across persons (experiencing positive outcomes does not necessarily mean that one might not also experience negative outcomes) the dominant reactions to these events, positive and negative affect, will be independent as well. On the other hand, if a response format is provided that asks subjects to indicate the *proportion of time* they had experienced positive or negative affect, the two would more likely be inversely correlated (Warr et al., 1983).

Following the publication of these three lines of work, as well as a related series of studies suggesting that separate personality dimensions, extraversion versus neuroticism, undergirded and perhaps caused the independence of positive and negative affect (Costa & McCrae, 1980; Emmons & Diener, 1986; Larsen, 1989; McCrae & Costa, 1983), a virtual cottage industry developed with the goal of demonstrating that positive and negative affect were indeed independent across a variety of contexts. Separate pleasant and unpleasant factors seemed to characterize the emotional experiences of bereaved individuals who had recently lost a spouse (Porritt & Bartrop, 1985). Independent positive and negative dimensions could be recovered in cross-cultural mood data, for example, in a large sample of Japanese subjects (Watson, Clark, & Tellegen, 1984). Relatively independent positive and negative affect factors were discovered in traditional, multidimensional mood scales like the Multiple Mood Adjective Checklist, although it was acknowledged that response sets may have contributed to the observed independence (Gotlib & Meyer, 1986). Data from adolescents who carried electronic beepers and who were asked to report on their moods at random times revealed frequency rates of positive and negative affect that were not correlated with each other (Larson, 1987). And, in perhaps the largest sample studied, Mayer and Gaschke (1988) confirmed that a two-dimensional structure of mood characterized the responses of nearly 1,600 undergraduates who completed three different mood scales (although, like Zevon and Tellegen [1982] before them, Mayer and Gaschke [1988] noted that bipolar Pleasant-Unpleasant and Arousal-Calm factors were simply rotated variants of the separate Positive Affect-Tired and Negative Affect-Relaxed factors). In recent years, the popularity of separate, orthogonal, positive or negative dimensions of mood has not waned, although the degree of independence is thought to vary as a function of the particular adjectives chosen (Watson, 1988), the intensity of the affects considered, and the time frame during which they were measured (Diener & Iran-Nejad, 1986), as well as the response format used to measure them (Warr et al., 1983).

Independence of Positive-Negative Affect: Systematic and Random Measurement Error

Although we acknowledge the potential theoretical value of considering positive and negative affect from an independent dimensions perspective (especially the importance for understanding the underpinnings of subjective well-being, see Diener, 1984, in press), it is the purpose of this article to call attention to methodological issues that may have caused the field prematurely to undervalue the traditional, bipolar view of affect.

In the study described earlier, Bentler (1969) noted that a systematic source of error variance, the acquiescent response style or the tendency to check adjectives of all kinds, masked the bipolar nature of semantic space, dramatically reducing correlations between the opposite ends of bipolar constructs. Writing more than 20 years ago, Bentler (1969) made the prescient observation that "rating scales in general are quite susceptible to an extremity response style. If an extremity response style existed. . . its effects would be to attenuate the potentially high negative correlation between polar oppositional terms. . . polar oppositional semantic tendencies might be negated by the existence of nonoppositional, or one-sided, irrelevant response tendencies" (pp. 34-35). Bentler was in fact identifying a particular exemplar of exactly the problem that we address here.

The purpose of the present set of studies is to demonstrate that questionnaire items about mood that are worded similarly and placed close together evoke similar response biases, giving rise to error covariation. When this source of nonrandom measurement error is combined with random error of measurement, correlations between positive and negative affect scales that bear no resemblance to true correlations may be generated. Constructs that are truly bipolar (large negative correlations between them) may appear to be independent (correlations near zero). Our hypothesis is that raw correlations and factor analyses that do not take the special properties of measurement error into account tend to suggest that positive and negative affect are largely independent. When systematic and random sources of error variance are accounted for, a bipolar model of affect emerges.

A Brief Look at How Measurement Error Distorts Measures of Association

When constructs are mismeasured, even strong underlying associations may turn up weak or incorrectly signed. Consider, by way of illustration, the product-moment correlation between happiness, denoted ξ_1, and sadness, ξ_2. If these latent constructs were observed without error, the correlation would be (i = individual)

$$r_{\xi_1\xi_2} = \frac{\mathrm{cov}(\xi_{1i}\xi_{2i})}{\sqrt{\quad}\sqrt{\quad}} \tag{1}$$

Suppose, however, that our information about these two variables is imperfect. Instead of observing ξ_{1i} and ξ_{2i}, we observe x_{1i} and x_{2i}, respectively:

$$X_{1i} = \xi_{1i} + \delta_{1i} \tag{2}$$

$$X_{2i} = \xi_{2i} + \delta_{2i}, \tag{3}$$

where δ_{1i} and δ_{2i} represent the error in measuring happiness and sadness, respectively.

Assume for the moment that the latent mood factors are independent of the measurement error terms and that the errors associated with the measure of happiness are independent of the errors associated with the measure of sadness. Thus, in the absence of sampling error,

$$\mathrm{COV}\,(\xi_{1i}, \xi_{2i}) = \mathrm{COV}\,(\xi_{1i}, \xi_{2i}) = 0 \qquad (4)$$

The correlation between X_{1i} and X_{21} would then be

$$\gamma_{x_1, x_2} = \frac{\mathrm{cov}\,(\xi_{1i}, \xi_{2i})}{\text{[illegible]}}. \qquad (5)$$

It follows directly that $|\gamma_{x_1, x_2}| < |\gamma_{\xi 1, \xi 2}|$ unless $\mathrm{var}\delta_{1i} = \mathrm{var}\delta_{2i} = 0$. In other words, random error produces correlation coefficients that are biased toward zero.

This, of course, is hardly news, especially to researchers working in the factor analytic tradition. The statistical problem, however, grows more complex when the data contain nonrandom response biases, whether due to acquiescence (Bentler,1969), extreme response style (Diener et al., 1985), or idiosyncratic use of response options (D. P. Green, 1988). In determining how systematic response biases contribute to the distortion of correlation coefficients, consider the following case. As before, we assume that the measurement errors are independent of the two moods, $cov(\xi_{ki}, \delta_{ki}) = 0$, but this time we allow response biases to affect both measures, such that $\mathrm{cov}(\delta_{1i}, \delta_{2i}) \neq 0$.

The correlation between x_{1i} and x_{2i} becomes

$$\gamma_{x_1, x_2} = \frac{\mathrm{cov}\,(\xi_{1i}, \xi_{2i}) + \mathrm{cov}\,(\delta_{1i}, \delta_{2i})}{\text{[illegible]}}. \qquad (6)$$

Depending on the sign and magnitude of $\mathrm{cov}(\delta_{1i}, \delta_{2i})$ and the reliability of the two proxies, $r_{x1,x2}$ may be greater than, less than, or equal to $\gamma_{\xi 1, \xi 2}$ In sum, random measurement error attenuates correlation coefficients; nonrandom error, however, may produce correlations that have the incorrect sign. Exploratory or confirmatory factor analyses of data collected using a single measurement approach often produce misleading results, unless special allowances are made for nonrandom error (Bank, Dishion, Skinner, & Patterson, 1990; Dillon, Kumar, & Mulani, 1987). The data presented below demonstrate that measurement error in single-method assessments of happy and sad mood states can lead one to overestimate their degree of independence. When multiple methods of data collection are used, confirmatory factor analysis indicates that happy and sad mood states are largely bipolar.

Study 1: The Bipolarity of Mood at Two Time Points

In the current study, we used a two-phase longitudinal design to examine the influence of response bias, both within and across time, on the dimensionality of mood. Data were collected at two time points 1 week apart. At each time, subjects completed four short affect measures. Each of the four affect measures differed in terms of response format and assessed both positive and negative mood. This multimethod approach to mood assessment allowed us to estimate the possible contribution of systematic method-specific influences on (a) the reliability of affect ratings and (b) the dimensional structure of mood.

To test the dimensional nature of mood, we performed confirmatory factor analysis (CFA) using both LISREL VI and VII (Joreskog & Sorbom, 1986,1988). This procedure finds the combination of parameters that maximizes the likelihood of obtaining the variance-covariance matrix of the observed sample data. CFA allows us to specify, a priori, the theoretical relations among latent mood factors, observed measures, and unique factors (also known as random errors). When constructs are measured with multiple indicators, it is possible to estimate both the interfactor correlations while simultaneously estimating intramethod correlations among the errors of measurement (Bollen, 1989). Thus, CFA allows us to consider both random and systematic variation when estimating the relations among the latent factors. Finally, CFA makes it possible to assess the statistical fit of one model in relation to other models that invoke different a priori assumptions.

Method

Subjects. Subjects were recruited from an introductory psychology course at a large northeastern university during the spring of 1991. Participation in the study was not a course requirement, no extra credit incentive was offered for experimental participation, and subjects who chose to participate did not receive financial compensation.

Of the 232 students who attended class during the first phase of data collection, 209 consented to participate in a study on mood. At the second data collection, one week later, 147 of the 209 participants returned questionnaires. However, because some items in either the first or second set of measures were left unanswered, complete information was not available for 8 subjects. As a result, the analyses presented here are based on 139 subjects (71 women and 68 men).

Procedure. The procedure for administering the questionnaires was similar at both data collection times. Before the start of class, subjects were asked to complete four mood measures to describe how they were feeling "this morning."[1] The mood measures were identical at each time, and the procedure took no more than 10 min.

Mood measures. There are three common response formats used to measure current emotional experiences: (a) the adjective checklist, (b) the response options format, and (c) the *n*-point Likert scale. These formats were used in designing four measures of mood for the current study: a mood adjective checklist, a response-options format that presented a list of statements to which the subjects indicated their degree of agreement by choosing a number ranging from 1 (*strong disagreement)* to 5 *(strong agreement),* a response-options format that presented statements to be rated from 1 (*very well)* to 4 (*not at all)* according to the degree to which they described the subject's mood, and a semantic differential Likert scale. Words and phrases used to create measures were drawn from lists of empirically derived synonyms (e.g., Izard, 1977; McNair et al., 1971) to represent either happy or sad mood. The actual terms used, as well as the question formats, are presented in Appendix A.

Results

Measurement model. The dimensionality of mood states is conventionally assessed by means of the correlation coefficient. As the relevant correlation is between positive and negative affect factors, rather than survey measures, statistical analysis requires a measurement model linking unobserved moods to observed indicators. The model we posit, which is presented in formal algebraic terms in Appendix B, supposes the observed score for each individual to be a linear combination of mood factor, systematic response bias, and random error. Note that our model does not assume that the data contain random

[1] Note that our measurement approach asks subjects to assess the general character of their moods, rather than the frequency with which they have experienced moods of different sorts (Larson, 1987).

and nonrandom error; if these measurement problems are not in evidence, the results will indicate this. The important assumptions of our model are that systematic response errors are uncorrected with mood factors and that measurement errors are uncorrelated across items with different question formats. These assumptions, like those of most any quantitative analysis, are not unassailable; but as we demonstrate below, our results are highly robust across a spectrum of different statistical assumptions.

An illustrative example. Let us begin, however, by disregarding the redundancy with which mood is measured in our study; for purposes of illustration, let us pretend we have only the 10 adjective checklist measures for happiness and sadness at one point in time. To meet the statistical identification requirements of CFA, we would create two subscales for happiness and for sadness. Arbitrarily arraying the mood adjectives in alphabetical order (the order in which they were presented), we would create one happiness scale by summing responses to the words *cheerful* and *elated;* another scale by adding responses to *glad, happy,* and *joyful.* Two corresponding sadness scales would be created in similar fashion. The correlation matrix of these four scales is listed in Table 1. Notice that on casual inspection, the matrix seems to suggest two mildly negatively correlated factors, one for the happy items and another for the sad. Indeed, when we posit a two-factor model that assumes only random error (cf. Long, 1983), that is what we find. The x^2 for this model is 0.01 *(df*= 1, N= 139, p = .92), and the estimated latent correlation between happiness and sadness is just -.34. Moreover, the statistical fit for the two-factor model completely dominates the fit associated with a nested one-factor alternative that assumes only random error is present. Computing the difference between the x^2s associated with the two models, a statistic that is itself distributed x^2, we obtain 65.26, *(df*= 1, N = 139, $p < .001$), which argues for a two-factor model.

The abysmal performance of the model that assumes a single, bipolar mood state seems to suggest that the mood states of happiness and sadness vary more or less independently of one another. But the one-factor model can be resuscitated by relaxing the assumption that the adjective checklist items contain only random error. The x^2 associated with the one-factor model can be driven from 65.27 to 0.14 if we assume that a consistent pattern of response bias (and hence covariance among the errors of measurement) runs throughout the adjective checklist. Note that even with this highly restrictive assumption— namely, each of the items absorbs the same amount of systematic measurement error variance—we can obtain a statistical fit that is on par with the two-factor model. In other words, fit statistics for these two nonnested alternative models may provide little guidance as to whether mood states are in fact bipolar (cf. Burke, Brief, George, Roberson, & Webster, 1989, p. 1095).

Table 1
Descriptive Statistics and Intercorrelations for Illustrative Example Described in Study 1

Ratings	1	2	3	4	*M*	*SD*
1. Happyl	—				0.39	0.60
2. Happy 2	.57**	—			0.48	0.72
3. Sad1	–.15*	–.18*	—		0.40	0.66
4. Sad2	–.21*	–.26**	.56**	—	0.39	0.71

Note. Happy1 = summed ratings for cheerful and elated; Happy2 = summed ratings for glad, happy, and joyful; Sad1 = summed ratings for blue and discouraged; Sad2 = summed ratings for feeling low, low, and sad.

* $p < .05$. ** $p < .001$.

The most sensible solution to this problem of indeterminacy would be to posit a pair of nested models that allows for varying degrees of nonrandom error while permitting two possibly distinct mood factors to emerge. The difficulty is that the parameters of these models will not be identified unless we have additional measures of mood—measures that are subject to different sorts of response bias. The essential feature of the present study is that it uses a variety of mood measures so as to overcome the identification problem. In short, using multiple measures, we develop a series of nested models that sustains the potential distinction between opposite mood states while also permitting nonrandom measurement error.

Correcting the observed correlation between happiness and sadness. The first and most parsimonious measurement model assumes that measurement error associated with each of the four types of survey questions is random. Each latent variable is assumed to take on the metric of the adjective checklist, leaving three unstandardized loadings to be estimated for each latent mood factor. In addition, we estimate one measurement error variance for each of the scales, for a total of 16 parameters. The four latent moods (happiness and sadness at both points in time) are assumed to be intercorrelated, adding another 10 parameters to be estimated. In all, the basic measurement model involves 38 parameters (12 factor loadings + 16 measurement error variances + 10 mood factor covariances).

Tables 2 and 3 offer a striking contrast between the raw interitem correlations and the interfactor correlations we obtain after correcting for random measurement error. Consider, for example, the correlation between the adjective checklist scale for happiness and that for sadness at the first point in time. Although this correlation is observed to be -.25 (see Table 2), the underlying correlation is estimated to be -.84 (see Table 3). This distortion is directly attributable to the poor reliability of the mood adjective checklist scales. The adjective checklist, however, is not the only scale containing measurement error. The strongest negative correlation observed in the first wave, —.69 between an *agree-disagree* scale and a Likert scale, nevertheless understates the actual degree of bipolarity of mood. Had we summed all our various mood measures into a single index, the observed correlation between happy and sad mood would have been -.72. Summing all measures other than the adjective checklist yields a correlation of-.74.

Mismeasurement need not be solely a matter of random error. Accordingly, the second model permits nonrandom error between survey items of similar response format. For example, the model allows for the possibility that a propensity to check boxes influenced individuals’ scores on each of the four adjective checklist scales. Similar allowance for response tendency was made for each of the other three measures, for a total of 24 additional parameters to be estimated.

Relaxing the assumption of random error produces a dramatic and statistically significant improvement in fit over the

Table 2
Raw Interitem Correlations Among All Indicators in Study 1

Indicator	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1. T1 H-ACL	–															
2. T1 H-A/D	.60	–														
3. T1 H-Desc	.49	.68	–													
4. T1 H-Likert	.55	.71	.67	–												
5. Tl S-ACL	–.25	–.49	–.51	–.50	–											
6. Tl S-A/D	–.39	–.64	–.59	–.69	.62	–										
7. T1 S-Desc	–.35	–.53	–.54	–.59	.64	.69	–									
8. T1 S-Likert	–.40	–.56	–.60	–.66	.56	.67	.74	–								
9. T2 H-ACL	.13	.10	.08	.05	.04	–.02 –	–.01	–.00	–							
10. T2 H-A/D	.00	.08	.13	.06	–.15	–.14	–.07	–.11	.60	–						
11. T2 H-Desc	–.01	.08	.21	.06	–.17	–.13	–.10	–.09	.55	.78	–					
12. T2 H-Likert	.02	.16	.20	.16	–.19	–.21	–.16	–.14	.63	.82	.79	–				
13. T2 S-ACL	.16	.01	–.13	.00	.21	.17	.18	.08	–.25	–.53	–.63	–.57	–			
14. T2 S-A/D	.01	.09	–.14	–.04	.18	.18	.19	.15	–.37	–.67	–.62	–.63	.61	–		
15. T2 S-Desc	.03	.02	–.10	.01	.11	.09	.19	.12	–.41	–.59	–.60	–.64	.62	.70	–	
16. T2 S-Likert	.11	.03	–.13	–.03	.14	.10	.16	.15	–.42	–.59	–.67	–.68	.61	.64	.69	–

Note. The letter before each format descriptor refers to the latent factor on which each indicator loads. Across-time correlations are enclosed in box. T1 = first data second data collection. H = happy; S = sad. ACL = Adjective Check List; A/D = agree-disagree format; Desc = descriptive statements format; Likert = Likert scale..

first model (difference in $x^2 = 76.5$, $df = 24$, $p < .001$, see Table 4). Other results, not reported in Table 4, are also of note. As we anticipated, the checklist scores are susceptible to nonrandom error. All six of the potential error correlations between checklist scales proved to be positive, as expected, and five were statistically significant using a one-tailed *t* test ($\alpha = .05$). There is evidence of nonrandom response effects among the other items as well. Three of the remaining 18 covariances between the errors of measurement proved to be significant at the .05 level (two-tailed), testifying to the existence of systematic response effects that occur within a survey and persist over time.[2]

Interestingly enough, the estimated interfactor correlations do not change much when our CFA model takes nonrandom error into account. This finding took us by surprise and led us to an interesting methodological insight: By measuring moods in a redundant fashion, one typically insulates a CFA analysis from the biasing effects of model misspecification. (The statistical basis for this conjecture can be found in a technical report by D. P Green, Goldman, & Salovey, 1992). In short, although assuming that the data contain only random error leads to the wrong CFA model, redundant measurement increases the number of elements in the covariance matrix that CFA analyzes that are free from the effects of nonrandom error. As a result, the CFA estimates from the "wrong" model approximate the true parameters. In the present case, our use of four different measures of each mood produces a covariance matrix of 120 offdiagonal elements, only 24 of which are contaminated by nonrandom error. Naturally, allowing for nonrandom error produces a better fitting statistical model, but the random error model does an adequate job of estimating the underlying parameters of interest.

Although relaxing the assumptions of our CFA model to incorporate both random and nonrandom error does not alter the estimated intermood correlations, it does change our explanation of why the observed correlations between adjective checklists are not stronger. For example, the observed correlation between adjective checklist measures of happiness and sadness in the second wave of our study is -.25. The corresponding latent correlation is estimated to be -.85 (random error model) or -.84 (nonrandom error model). Our estimates based on a CFA model that allows for both random and nonrandom error suggest that random error drives the latent correlation down to -.42; nonrandom error reduces it further to -.29.[3]

Lest one think that CFA transforms all weak correlations into strong ones, we find that the over-time stability of the four mood states is quite weak. The latent degree of sadness at Time 1 displays a correlation of just .20 with sadness 1 week later, and the other over-time correlations are even weaker. The finding is of substantive interest because it suggests that moods are not enduring characteristics, even over short periods of time. This finding has important implications for interpreting responses to questions that ask subjects to report their moods over the past month or year. If moods are highly transitory, such measures may be better viewed as tapping mood tendencies or mood predispositions.

[2] Had we only allowed for nonrandom error between common formats within a given wave, thereby ignoring the problem of response biases that persist over time, the result would have been a significant improvement in fit ($p < .001$) over the random error model but no change in the estimated interfactor correlations.

[3] The slight difference between this correlation and the actual observed correlation is an unexplained residual attributed to sampling variability.

Table 3

Factor Loadings and Interfactor Correlations for Random Error Model at Time 1 and Time 2 in Study 1

Indicator	T1 Happy	T1 Sad	T2 Happy	T2 Sad
Standardized factor loadings[a]				
H-ACL	.63		.66	
H-A/D	.78		.87	
H-Desc	.86		.92	
H-Likert	.85		.89	
S-ACL		.72		.75
S-A/D		.85		.83
S-Desc		.83		.81
S-Likert		.83		.83
Interfactor correlations				
T1 happy	–			
T1 sad	–.84**	–		
T2 happy	.16	–.18	–	
T2 sad	–.02	.23*	–.85**	—

Note. The letter before each format descriptor refers to the latent factor on which each indicator loads. H = happy; S = sad; T = time; ACL = Adjective Check List; A/D = agree-disagree format; Desc = descriptive format; Likert = Likert scale.
[a]X^2(98, *N* = 139) = 162.39, *p* < .001. Goodness of Fit Index = .87. Adjusted Goodness of Fit Index = .83. Root-mean-squared residual = .05.
* *p* < . 05. ***p*<.001.

Factor Loadings and Interfactor Correlations for Correlated Error Model at Time 1 and Time 2 in Study 1

Indicator	T1 Happy	T1 Sad	T2 Happy	T2 Sad
Standardized factor loadings[a]				
H-ACL	.65		.69	
H-A/D	.80		.85	
H-Desc	.84		.92	
H-Likert	.86		.89	
S-ACL		.71		.72
S-A/D		.86		.85
S-Desc		.83		.82
S-Likert		.83		.80
Interfactor correlations				
T1 happy	–			
T1 sad	–.84**	–		
T2 happy	.14	–.17	–	
T2 sad	–.04	.20*	–.84**	–

Note. The letter before each format descriptor refers to the latent factor on which each indicator loads. H = happy; S = sad; T = time; ACL = Adjective Check List; A/D = agree-disagree format; Desc = descriptive format; Likert = Likert scale.
[a] x^2(74, *N*= 139) = 85.90, *p* = *ns.* Goodness of Fit Index = .93. Adjusted Goodness of Fit Index = .87. root-mean-squared residual = .05. * *p* < .05.

Study 2: Replication Using Varied Question Order

During the fall of 1991, we replicated our initial study using a sample of students enrolled in an undergraduate psychology course at the same university. The administration of the study was the same, except that the sequence of the different formats for mood assessment was varied according to a Latin-square design. Of the 320 students in the class, 285 completed mood-assessment surveys. Eliminating respondents who offered invalid or missing answers to any of the items left 250 valid cases.

The second study confirmed the main conclusions of our initial CFA analysis: Happiness and sadness appear to be bipolar in structure, and appearances to the contrary may be attributed to random and nonrandom error (Tables 5 and 6). The observed correlation between the happy and sad mood adjective scales is -.40. When we take random error into account, this correlation jumps to -.92. Allowing for nonrandom response errors between items of similar question wording and response format cut the x^2 from 75.20 to 29.75 with an expenditure of just four degrees of freedom *(N* = 250, *p* < .001). Again, amid strong evidence of nonrandom error, there was little change in the estimated correlation between happy and sad mood states as we relaxed the constraints of the CFA model (*r*= —.91).

Study 3: Replication Using Different Time Horizon and Adjectives

Five weeks after the administration of Study 2, we replicated that study using the same pool of subjects. This time, however, the time frame of our mood assessment questions was changed from "how you have been feeling since this morning" to "how you have been feeling over the past month" so as to test whether a longer time horizon diminishes the bipolarity of mood observed above (as suggested by Diener & Iran-Nejad, 1986). Again, the sequence of the different types of mood assessment was varied according to a Latin-square design. Because the questionnaire was administered close to the end of the semester, attendance was up, and 304 students completed mood assessment surveys.

The data provided little support for the notion that the bipolarity of mood is a function of the time frame of the mood assessment (Tables 7 and 8). A CFA model assuming random measurement error produced a disattenuated correlation of

Table 5

Raw Interitem Correlations for Happy and Sad Words: Study 2

Indicator	1	2	3	4	5	6	7	8
1. H-ACL	—							
2. H-A/D	.65	–						
3. H-Desc	.63	.71	–					
4. H-Likert	.67	.76	.73	–				
5. S-ACL	–.40	–.62	–.58	–.57	–			
6. S-A/D	–.50	–.66	–.58	–.66	.67	–		
7. S-Desc	–.55	–.59	–.69	–.60	.54	.55	–	
8. S-Likert	–.61	–.68	–.68	–.75	.65	.76	.60	—

Note. The letter before each format descriptor refers to the latent factor on which each indicator loads. H = happy; S = sad; ACL = Adjective Check List; A/D = agree-disagree format; Desc = descriptive format; Likert = Likert scale.

Table 6
Standardized Factor Loadings and Interfactor Correlations for Random and Correlated Error Models for Happy and Sad Words: Study 2

Model and indicator	Factor loadings	
	Happy	Sad
Random error[a]		
H-ACL	.74	
H-A/D	.86	
H-Desc	.83	
H-Likert	.89	
S-ACL		.75
S-A/D		.83
S/Desc		.71
S-Likert		.89
Correlated error[b]		
H-ACL	.75	
H-A/D	.87	
H-Desc	.82	
H-Likert	.87	
S-ACL		.76
S-A/D		.84
S-Desc		.70
S-Likert		.88

Note. The letter before each format descriptor refers to the latent factor on which each indicator loads. H = happy; S = sad; ACL = Adjective Check List; A/D = agree-disagree; Desc = descriptive format; Likert = Likert scale.
[a] $X^2(19, N = 250) = 75.20$, $p < .001$; Goodness of Fit Index = .93; Adjusted Goodness of Fit Index = .86; RMSR = .04. Happy-sad interfactor correlation = -.92 ($p < .001$).
[b] $x^2(15, N = 250) = 29.75$, $p = .025$; Goodness of Fit Index = .97; Adjusted Goodness of Fit Index = .93; RMSR = .03. Happy-sad interfactor correlation = -.91 ($p < .001$).

Table 8
Standardized Factor Loadings and Interfactor Correlations for Random and Correlated Error Models for Happy and Sad Words: Study 3

Model and indicator	Factor loadings	
	Happy	Sad
Random error[a]		
I. H-ACL	.67	
2. H-A/D	.79	
3. H-Desc	.87	
4. H-Likert	.86	
5. S-ACL		.72
6. S-A/D		.85
7. S-Desc		.73
8. S-Likert		.84
Correlated error[b]		
1. H-ACL	.67	
2. H-A/D	.80	
3. H-Desc	.86	
4. H-Likert	.85	
5. S-ACL		.72
6. S-A/D		.85
7. S-Desc		.72
8. S-Likert		.82

Note. The letter before each format descriptor refers to the latent factor on which each indicator loads. H = happy; S = sad; ACL = Adjective Check List; A/D = agree-disagree format; desc = descriptive format; Likert = Likert scale.
[a] $x^2(19, N = 304) = 135.50$, $p < .001$; Goodness of Fit Index = .91; Adjusted Goodness of Fit Index = .83; root-mean-squared residual = .06. Happy-sad interfactor correlation = -.87 ($p < .001$).
[b] $X^2(15, N = 304) = 25.54$, $p = .04$; Goodness of Fit Index = .98; Adjusted Goodness of Fit Index = .95; root-mean-squared residual = .03. Happy-sad interfactor correlation = -.87 ($p < .001$).

-.86 between happiness and sadness. Allowance for both random and nonrandom error increased this estimate to -.87.

In keeping with previous findings, CFA assuming random and nonrandom error produced similar substantive conclusions. Again, however, there can be no doubt about the prevalence of nonrandom error in these mood assessments. The estimated covariance between response errors associated with the two mood checklist scales is more than 8 times its standard

Table 7
Raw Interitem Correlations for Happy and Sad Words: Study 3

Indicator	1	2	3	4	5	6	7	8
1. H–ACL	—							
2. H–A/D	.60	—						
3. H–Desc	.59	.69	—					
4. H–Likert	.59	.67	.74	—				
5. S–ACL	–.10	–.46	–.53	–.50	—			
6. S–A/D	–.44	–.53	–.63	–.64	.68	—		
7. S–Desc	–.44	–.48	–.59	–.59	.52	.60	—	
8. S–Likert	–.47	–.59	–.65	–.66	.60	.69	.61	—

Note. The letter before each format descriptor refers to the latent factor on which each indicator loads. H = happy; S = sad; ACL = Adjective Check List; A/D = agree-disagree format; Desc = descriptive format; Likert = Likert scale.

error, suggesting that respondents have markedly different propensities to check off adjectives whether due to individual differences in acquiescence, expressiveness, or emotional intensity (Bentler, 1969; Diener, Larsen, Levine, & Emmons, 1985). Viewing the three studies together, we found that all four of the within-wave correlations between the checklist measures were significantly inflated by nonrandom error ($p < .05$). Like Bentler (1969) before us, we advise caution when researchers analyze data obtained with a checklist format.[4]

Results for the other formats suggest that they vary in their susceptibility to nonrandom error. The Likert scales produced negatively correlated measurement errors in all four cases, two significantly so. The same pattern of results obtained for the agree-disagree format. The self-description items, however, tended to produce positive error correlations (3 of 4), although only one proved significantly positive. One feature that distinguished the self-description items from the Likert scales and agree-disagree format is that the latter offered subjects a non-

[4] A PsycLIT scan for the years 1987-1991 listed more than 25 abstracts that mentioned the mood adjective checklist as the primary method by which mood was assessed. In addition, a recently developed mood measure uses an adjective checklist format (Matthews, Jones, & Chamberlain, 1990).

committal middle response option. As D. P. Green (1988) has argued, negative error covariance is typical of items with attractive middle alternatives.

The strongest case for the independence of positive and negative mood has been advanced by Watson, Clark, and Tellegen (1988) using adjectives drawn from the mood dimensions formed by the 45° rotation described in the introduction and labeled *Positive Affect* and *Negative Affect.* Positive affect defined in this fashion comprises moods such as excited or enthusiastic rather than happy; negative affect refers to moods such as distressed or nervous rather than sad. Because both mood states operate in a state of arousal, we would not expect the latent dimensions to bear a perfect negative correlation with one another. In essence, arousal represents a third, oblique factor related positively to both positive and negative affect. The question of interest is whether positive and negative affect are orthogonal, as a two-dimensional exploratory factor analytic solution of the Positive and Negative Affect Schedule (the PANAS) would suggest (Watson, Clark, & Tellegen, 1988).

To address this question, we included in the questionnaire administered in Study 3 a series of mood items derived from the PANAS. Again, four formats were used to measure mood. The time frame specified was "the past month," so as to give the independence hypothesis its best opportunity to acquit itself. The words used to assess positive and negative affect are listed in Appendix C. Our statistical analysis is based on 305 subjects with valid responses.

Our analysis offers no support for the view that positive and negative affect, operationalized using the PANAS adjectives, represent orthogonal dimensions (see Tables 9 and 10). The observed correlations between the two mood states range from -.11 to -.47, and correction for measurement error places the estimated interfactor correlation at -.57 (random error) or-.58 (nonrandom error). A formal statistical test of the hypothesis that the two factors are orthogonal yields an unequivocal answer: A x^2 difference test with 1 degree of freedom comparing the nonrandom error model with a similar model in which the interfactor correlation is constrained to be zero produces a value of 81.23, $p < .001$. Positive and negative affect, although perhaps not as bipolar as happy and sad moods, cannot be regarded as orthogonal factors.

Table 9
Raw Interitem Correlations for Positive and Negative Affect Words: Study 3

Indicator	1	2	3	4	5	6	7	8
1. P–ACL	—							
2. P–A/D	.44	—						
3. P–Desc	.62	.55	—					
4. P–Likert	.65	.54	.64	—				
5. N–ACL	–.11	–.20	–.23	–.14	—			
6. N–A/D	–.25	–.27	–.30	–.37	.50	—		
7. N–Desc	–.35	–.31	–.45	–.44	.46	.61	—	
8. P–Likert	–.31	–.31	–.40	–.47	.52	.61	.61	—

Note. The letter before each format descriptor refers to the latent factor on which each indicator loads. P = positive affect; N = negative affect; ACL = Adjective Check List; A/D = agree-disagree format; Desc = descriptive format; Likert = Likert scale.

Standardized Factor Loadings and Interfactor Correlations for Random and Correlated Error Models for Positive and Negative Words: Study 3

Model and indicator	Factor loading: Positive	Negative
Random error[a]		
P-ACL	.75	
P-A/D	.65	
P-Desc	.80	
P-Likert	.84	
N-ACL		.60
N-A/D		.77
N-Desc		.79
N-Likert		.80
Correlated error[b]		
P-ACL	.71	
P-A/D	.65	
P-Desc	.81	
P-Likert	.80	
N-ACL		.60
N-A/D		.77
N-Desc		.78
N-Likert		.79

Note. The letter before each format descriptor refers to the latent factor on which each indicator loads. P = positive; N = negative; ACL= Adjective Check List; A/D = agree-disagree format; desc = descriptive format; Likert = Likert scale.
[a] $X^2(19, N = 305) = 124.09$, $p < .001$. Goodness of Fit Index = .91. Adjusted Goodness of Fit Index = .84. root-mean-squared residual = .07. Positive-negative interfactor correlation = -.57 ($p < .001$).
[b] $X^2(15, N = 305) = 41.59$, $p < .001$. Goodness of Fit Index = .97. Adjusted Goodness of Fit Index = .92. root-mean-squared residual = .04. Positive-negative interfactor correlation = -.58.

General Discussion

The findings presented in these studies underscore the importance of multimethod research designs (Campbell & Fiske, 1959). Although this methodological principle guides research in some areas of social science (cf. Bank, Dishion, Skinner, & Patterson, 1990), it is often honored in the breach. This is particularly true of mood research, where multimethod designs are seldom used to assess the structure of affective states. As a result, the empirical analysis of mood has been prone to statistical artifacts, such as the finding that positive and negative moods are weakly correlated. When random and nonrandom sources of error are taken into account using multiple methods of mood assessment, a largely bipolar structure for affect emerges.[5]

Previous researchers in this area have not been oblivious to the problem of measurement error. Most, in fact, have taken pains to assess the reliability of the mood adjective checklist

[5] Although the negative interfactor correlations we obtain between pleasant and unpleasant moods are extremely high, they are not -1.0. In none of our studies does a nested x^2 difference test enable us to accept the null hypothesis that one rather than two factors generated the data. Following Bentler (1969), we dub these mood states "approximately" bipolar.

scales. The problem is that without redundancy in method, one cannot readily assess reliability. Reliability assessment as it is usually performed assumes random measurement error, but additive scales, particularly those constructed from adjective checklists, may well contain systematic response bias as well. Because nonrandom error can lead to inflated reliability estimates (D. P Green & Citrin, in press), researchers are led to believe that observed correlations are less attenuated than they really are, which only reinforces the tendency to conclude that opposite affective states are relatively independent.[6]

Another drawback of the single measurement approach is that it can throw off assessments of discriminant validity. However tempting it might be to regress some theoretically telling dependent variable on pleasant and unpleasant mood scales, random and nonrandom error are likely to produce misleading results. As Achen (1985) has demonstrated, the coefficients from this regression will tend to be biased, sometimes severely. Thus, even if happiness and sadness were perfectly bipolar, ordinary least squares regression might suggest that both have independent predictive effects of different magnitudes. A review of studies demonstrating that positive and negative moods are differentially related to various criteria—traits, life satisfaction, and clinical diagnoses—lies beyond the scope of this article. We should note, however, that this sort of validity research brings its own set of statistical complications to the study of mood structure.

Perhaps the strongest argument in favor of a multimethod approach is that it greatly enhances the robustness of statistical tests of bipolarity. Despite strong evidence that our data are contaminated with nonrandom error, the precise way in which measurement error is modeled has little bearing on the substantive conclusions generated by CFA. Across a variety of model specifications, the underlying structure of pleasant and unpleasant mood states turns out to be approximately bipolar, at least for the short to medium time frames measured here. As we noted in the introduction, this is by no means a new conclusion. A vast number of investigators, however, operate under the opposite assumption. A cursory scan of recent work using PsycLIT netted no fewer than 193 empirical studies between 1987 and 1991 that measured positive and negative mood as distinct dimensions. In addition, many studies and clinical practices have operated under the assumption that positive affect and negative affect, as defined by Watson (1988) and measured by the PANAS adjectives, are not only distinct, but orthogonal dimensions. Although we do not contend that these constructs are perfectly bipolar, the correlation between them, net of measurement error, is substantial.

Beyond these substantive conclusions, however, lie some practical insights about how one might develop more effective measures of mood states. An alternative to Bentler's (1969) method of administering an adjective checklist with a large number of items (with which to "partial out" the effects of nonrandom error) is a series of short question batteries of different format. Granted, the adjective checklist scale will contain more random noise when the number of adjectives is reduced, but the abundance of across-method correlations makes reliability correction possible and, as we have seen, robust.

Although the multimethod design we are proposing is relatively modest in comparison with the ambitious multisource research design proposed by Bank et al. (1990), it is clearly at odds with current practice. It is conventional to ask subjects to respond to many items of the same type so as to streamline the survey instrument and minimize the amount of instructions that must be given to the respondent. In our view, the benefits of using a one-format questionnaire are small in comparison with the costs of potential biases due to systematic response effects. Three methodologically distinct 2-item assessments of mood states, for example, are likely to be much more statistically informative than a single 20-item assessment.[7] The latter, after all, provides limited opportunity to gauge and counteract the effects of systematic response bias.

The methodological argument we have advanced extends beyond the scope of mood research. Consider, for example, a topic that has puzzled political scientists since the early 1980s: the structure of public sentiment toward partisan and ideological groups. Intuition suggests that people who like liberals would tend to dislike conservatives; those who like Democrats dislike Republicans. But at first glance, the data suggest otherwise. The weak zero-order correlations between these evaluative dimensions seem to indicate that people who feel warmly toward conservatives do not tend to dislike liberals (Conover & Feldman, 1981). Similarly, survey data suggest that feelings about Democrats are relatively independent of feelings about Republicans (Weisberg, 1980). Yet, when these findings are reevaluated using a multimethod CFA, random and nonrandom measurement error conceal underlying correlations of-.80 and up (D. P. Green, 1988).

Perhaps the best illustration of how measurement error can distort the strength of negative associations comes from the National Election Survey of 1964, in which respondents were asked to rate the "warmth" they felt toward a variety of political

[6] A good illustration of how standard approaches to reliability enhancement fail to overcome problems of nonrandom error is presented in the Bentler (1969) study: His scales comprised dozens of items and therefore appeared highly reliable. Yet, because his mood scales were fraught with nonrandom error, he obtained weak bivariate correlations between opposite mood states. Bentler was able to correct this problem by devising another highly reliable measure of the response bias itself and using this measure to "partial out" the positive correlation among the measurement errors. The techniques described here achieve the same result but do not require the investigator to administer a prodigious mood adjective checklist inventory.

[7] Consider, for example, two alternative ways of studying the dimensionality of positive and negative affect, as defined by Watson (1988). One approach (see Study 3) would be to use three sets of two-item indices (semantic differential, self-description, and agree-disagree). This design takes only minutes to administer, facilitates a two-factor model that allows for nonrandom error with 5 degrees of freedom, and suggests an interfactor correlation of-.58. Another approach would be to administer only the original 20-item PANAS scale. This battery takes about the same amount of time to complete but does not provide the data analyst any latitude in the way measurement error may be modeled (cf. Green & Citrin, in press). Although with appropriate correction for mismeasurement, both approaches yield the same answer, the former presents substantially fewer modeling problems. Moreover, data collected in a single-method fashion give little diagnostic information that might alert the researcher to the problems created by mismeasurement.

and social groups on a "feeling thermometer" ranging from 0 to 100. Two well-known groups on this list were the Ku Klux Klan (KKK) and the National Association for the Advancement of Colored People (NAACP). It is hard to imagine that people who felt warmly toward the NAACP in 1964 could feel anything but contempt for the KKK and vice versa, yet the correlation between these two items turned out to be only -.31 (N= 1,454). Far from indicating the multidimensional character of public sentiment toward these two groups, the data attest to the low reliability and systematic response biases that plague this type of survey measure (D. P. Green, 1988).

Anomalous results of this kind should direct our attention back to the way in which the data were generated. Were the data collected using one or many methods? How reliable are the measures? The point is not that two-factor solutions should be rejected out of hand, but rather that this type of research finding requires special methodological attention. Although the present analysis does not resolve the debate about the structure of mood, we do provide evidence to suggest that claims of independence deserve a second look. Until the issue is resolved, investigators would best be advised to measure mood using several different methods within the same study (i.e., varied question and response formats, and if feasible, data obtained through psychophysical assessment or clinical observation), even if this means sacrificing the extensiveness of any single battery of measures.

References

Achen, C. H. (1985). Proxy variables and incorrect signs on regression coefficients. *Political Methodology, 11,* 299–316.

Bank, L., Dishion, T,, Skinner, M,, & Patterson, G. R. (1990). Method variance in structural equation modeling: Living with "glop." In G. R. Patterson (Ed.), *Depression and aggression in family interaction* (pp. 247–279). Hillsdale, NJ: Erlbaum.

Bentler, P. M. (1966). *Multidimensional homogeneity scaling.* Paper presented at the Special Meeting of the Psychometric Society, New York.

Bentler, P. M. (1969). Semantic space is (approximately) bipolar. *Journal of Psychology, 71,* 33–40.

Bollen, K. A. (1989). *Structural equations with latent variables.* New York: Wiley.

Borgatta, E. F. (1961). Mood, personality and interaction. *Journal of General Psychology, 64,* 105–137.

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

Bradburn, N. M., & Caplovitz, D. (1965). *Reports on happiness.* Chicago: Aldine.

Burke, M. J., Brief, A. P., George, J. M., Roberson, L., & Webster, J. (1989). Measuring affect at work: Confirmatory factor analyses of competing mood structures with conceptual linkage to cortical regulatory systems. *Journal of Personality and Social Psychology, 57,* 1091–1102.

Campbell, D. T., & Fiske, D. W (1959). Convergent and discriminant validation by the multitrait–multimethod matrix. *Psychological Bulletin, 56,* 81–105.

Clyde, D. J. (1963). *Manual for the Clyde Mood Scale.* Coral Gables, FL: Biometric Laboratory, University of Miami.

Conover, P., & Feldman, S. (1981). The origins and meaning of liberal–conservative self–identifications. *American Journal of Political Science, 25,* 617–645.

Costa, P. T., & McCrae, R. R. (1980). Influence of extraversion and neuroticism on subjective well–being: Happy and unhappy people. *Journal of Personality and Social Psychology, 38,* 668–678.

Diener, E. (1984). Subjective well–being. *Psychological Bulletin, 95,* 542–575.

Diener, E. (in press). Challenges in measuring subjective well–being and ill–being. *Psychological Bulletin.*

Diener, E., & Emmons, R. A. (1984). The independence of positive and negative affect. *Journal of Personality and Social Psychology, 47,* 1105–1117.

Diener, E., & Iran–Nejad, A. (1986). The relationship in experience between various types of affect. *Journal of Personality and Social Psychology, 50,* 1031–1038.

Diener, E., Larsen, R. J., Levine, S., & Emmons R. A. (1985). Intensity and frequency: Dimensions underlying positive and negative affect. *Journal of Personality and Social Psychology, 48,*1253–1265.

Dillon, W. R., Kumar, A., & Mulani, N. (1987). Offending estimates in covariance structure analysis: Comments on the causes of and solutions to Heywood cases. *Psychological Bulletin, 101,*126–135.

Emmons, R. A., & Diener, E. (1986). Influence of impulsivity and sociability on subjective well–being. *Journal of Personality and Social Psychology, 50,* 1211–1215.

Gotlib, I. H., & Meyer, J. P. (1986). Factor analysis of the multiple affect adjective check list: A separation of positive and negative affect. *Journal of Personality and Social Psychology, 50,* 1161–1165.

Green, D. P. (1988). On the dimensionality of public sentiment toward partisan and ideological groups. *American Journal of Political Science, 32,* 758–780.

Green, D. P., & Citrin, J. (in press). Measurement error and the structure of attitudes: Are positive and negative judgments opposites? *American Journal of Political Science*

Green, D. P., Goldman, S. L., & Salovey, P. (1992). *The effects of nonrandom error on confirmatory factor analysis: How multiple–methods improve robustness.* Unpublished manuscript, Yale University, New Haven, CT.

Green, R. F., & Goldfried, M. R. (1965). On the bipolarity of semantic space. *Psychological Monographs, 79* (Whole No. 599).

Guilford, J. P. (1954). *Psychometric methods* (2nd ed). New York: McGraw–Hill.

Izard, C. E. (1972). *Patterns of emotions.* San Diego, CA: Academic Press.

Izard, C. E. (1977). *Human emotions.* New York: Plenum Press.

Jöreskog, K. G, & Sörbom, D. (1986). *LISREL VI.* Mooresville, IN: Scientific Software.

Jöreskog, K. G,, & Sörbom, D. (1988). *LISREL VII.* Mooresville, IN: Scientific Software.

Larsen, R. J. (1989, August). *Personality as an affect dispositional system.* Paper presented at the 97th Annual Convention of the American Psychological Association, New Orleans, LA.

Larsen, R. J., & Diener, E. (1992). Promises and problems with the circumplex model of emotion. *Review of Personality and Social Psychology, 13,* 25–59.

Larson, R. W (1987). On the independence of positive and negative affect within hour–to–hour experience. *Motivation and Emotion, 11,* 145–156.

Long, J. S. (1983). *Confirmatory factor analysis.* Sage University Paper series on Quantitative Applications in the Social Sciences (Series No. 07–033). Beverly Hills, CA: Sage.

Matthews, G., Jones, D. M., & Chamberlain, A. G. (1990). Refining the measurement of mood: The UWIST mood adjective checklist. *British Journal of Psychology, 81,* 17–42.

Mayer, J. D., & Gaschke, Y. N. (1988). The experience and meta–experience of mood. *Journal of Personality and Social Psychology, 55,* 102–111.

McCrae, R. R., & Costa, P. T. (1983). Psychological maturity and sub-

jective well–being: Toward a new synthesis. *Developmental Psychology, 19,* 243–248.
McNair, D. M., & Lorr, N. (1964). An analysis of mood in neurotics. *Journal of Abnormal and Social Psychology, 69,* 620–627.
McNair, D. M., Lorr, M., & Droppleman, L. F. (1971). *Manual: Profile of mood states.* San Diego, CA: Educational and Industrial Testing Service.
Meddis, R. (1972). Bipolar factors in mood adjective checklists. *British Journal of Social and Clinical Psychology, 11,*178–184.
Moore, B. S., & Isen, A. M. (1990). Affect and social behavior. In B. S. Moore & A. M. Isen (Eds.), *Affect and social behavior* (pp. 1–21). Cambridge, England: Cambridge University Press.
Nowlis, V (1965). Research with the Mood Adjective Check List. In S. S. Tomkins & C. E. Izard (Eds.), *Affect, cognition, and personality* (pp. 352–389). New York: Springer.
Nowlis, V, & Nowlis, H. H. (1956). The description and analysis of mood. *Annals of the New York Academy of Sciences, 65,* 345–355.
Osgood, C., Suci, G., & Tannenbaum, P. (1957). *The measurement of meaning.* Urbana: University of Illinois Press.
Porritt, D., & Bartrop, R. W. (1985). The independence of pleasant and unpleasant affect: The effects of bereavement. *Australian Journal of Psychology, 37,* 205–213.
Russell, J. A. (1978). Evidence of convergent validity on the dimensions of affect. *Journal of Personality and Social Psychology, 36,* 1152–1168.
Russell, J. A. (1979). Affective space is bipolar. *Journal of Personality and Social Psychology, 37,* 1161–1178.
Russell, J. A., & Mehrabian, A. (1977). Evidence for a three–factor theory of emotions. *Journal of Research in Personality, 11,* 273–294.
Thayer, R. E. (1967). Measurement of activation through self–report. *Psychological Reports, 20,* 663–678.
Warr, P., Barter, J., & Brownbridge, G. (1983). On the independence of positive and negative affect. *Journal of Personality and Social Psychology, 44,* 644–651.
Watson, D. (1988). The vicissitudes of mood measurement: Effects of varying descriptors, time frames, and response formats on measures of positive and negative affect. *Journal of Personality and Social Psychology, 55,* 128–141.
Watson, D., Clark, L. A., & Tellegen, A. (1984). Cross–cultural convergence in the structure of mood: A Japanese replication and a comparison with U.S. findings. *Journal of Personality and Social Psychology, 47,* 127–144.
Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of a brief measure of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54,*1063–1070.
Watson, D., & Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological Bulletin, 98,* 219–235.
Weisberg, H. F. (1980). A multidimensional conceptualization of party identification. *Political Behavior, 2,* 33–60.
Wundt, W. (1897). *Outlines of psychology*(C. H. Judd, Trans.). Leipzig, Germany, Wilhelm Englemann.
Zevon, M. A., & Tellegen, A. (1982). The structure of mood change: An idiographic/nomothetic analysis. *Journal of Personality and Social Psychology, 43,* 111–122.

Appendix A

Items Used to Measure Happy and Sad Moods in Studies 1–3

Study 1

Adjective Checklist

Cheerful	*Blue*
Elated	*Discouraged*
Glad	*Feeling low*
Happy	*Low*
Joyful	*Sad*

Agree-Disagree Response Option (5 points)

I have been in a cheerful mood. All in all, I’ve been feeling despondent.

“Describes Me” Response Option (4 points)

I have been in a good mood. I have felt sad and dispirited.

Unipolar Likert Scale (7 points)

Happy-not happy
Discouraged-not discouraged

Studies 2 and 3

Adjective Checklist

Cheerful	*Blue*
Contented	*Depressed*
Happy	*Downhearted*
Pleased	*Gloomy*
Satisfied	*Sad*
Warmhearted	*Unhappy*

Agree-Disagree Response Option (5 points)

I’ve been in good spirits.
All in all, I’ve been feeling kind of depressed.

“Describes Me” Response Option (4 points)

I have been in a good mood.
I’ve felt sad and dispirited.

Unipolar Likert Scale (7 points)

Happy-not happy
Discouraged-not discouraged

Appendix B

Table B1

Measurement Equations for Confirmatory Factor Analyses

Factor	Random error	Nonrandom error
ξ_1 = positive affect factor, Time 1 =	$x_1 = \lambda_{11}\xi_1 + \delta_1$	$x_1 = \lambda_{11}\xi_1 + \delta_{1A}$
	$x_2 = \lambda_{21}\xi_1 + \delta_2$	$x_2 = \lambda_{21}\xi_1 + \delta_{2B}$
	$x_3 = \lambda_{31}\xi_1 + \delta_3$	$x_4 = \lambda_{31}\xi_1 + \delta_{3C}$
	$x_4 = \lambda_{41}\xi_1 + \delta_4$	$x_5 = \lambda_{41}\xi_1 + \delta_{4D}$
ξ_2 = negative affect factor, Time 1 =	$x_5 = \lambda_{52}\xi_2 + \delta_5$	$x_5 = \lambda_{52}\xi_2 + \delta_{5A}$
	$x_6 = \lambda_{62}\xi_2 + \delta_6$	$x_6 = \lambda_{62}\xi_2 + \delta_{6B}$
	$x_7 = \lambda_{72}\xi_2 + \delta_7$	$x_7 = \lambda_{72}\xi_2 + \delta_{7C}$
	$x_8 = \lambda_{82}\xi_2 + \delta_8$	$x_8 = \lambda_{82}\xi_2 + \delta_{8D}$
ξ_3 = positive affect factor, Time 2 =	$x_9 = \lambda_{93}\xi_3 + \delta_9$	$x_9 = \lambda_{93}\xi_3 + \delta_{9A}$
	$x_{10} = \lambda_{10,3}\xi_3 + \delta_{10}$	$x_{10} = \lambda_{10,3}\xi_3 + \delta_{10B}$
	$x_{11} = \lambda_{11,3}\xi_3 + \delta_{11}$	$x_{11} = \lambda_{11,3}\xi_3 + \delta_{11C}$
	$x_{12} = \lambda_{12,3}\xi_3 + \delta_{12}$	$x_{12} = \lambda_{12,3}\xi_3 + \delta_{12D}$
ξ_4 = negative affect factor, Time 2 =	$x_{12} = \lambda_{13,4}\xi_4 + \delta_{13}$	$x_{13} = \lambda_{13,4}\xi_4 + \delta_{13A}$
	$x_{12} = \lambda_{14,4}\xi_4 + \delta_{14}$	$x_{14} = \lambda_{14,4}\xi_4 + \delta_{14B}$
	$x_{12} = \lambda_{15,4}\xi_4 + \delta_{15}$	$x_{15} = \lambda_{15,4}\xi_4 + \delta_{15C}$
	$x_{12} = \lambda_{16,4}\xi_4 + \delta_{16}$	$x_{16} = \lambda_{16,4}\xi_4 + \delta_{16D}$

Note. The measures x_1, x_5, x_9, and x_{13} represent mood adjective checklist scales. Similarly, the x_2, x_6, x_{10}, and x_{14} represent agree-disagree items; x_3, x_7, x_{11}, and x_{15} are self-description items; and x_4, x_8, x_{12}, and x_{16} are semantic differential Likert scales. The λs are factor loadings, and the δs are individual elements in the error matrix. All measures depicted within braces are assumed to load on the mood factor named to the left. The λ_k associated with the mood adjective checklist measures are assumed to be unity, to set the metric for the latent factors. The four factor variances and six interfactor covariances (ϕ_{ij}) are free parameters. The 16 measurement error variances (θ^{δ}_{kk}) are free parameters. Analyses in which allowance is made for nonrandom error assume that all δ_ks ending in the same subscript letter are potentially correlated. This adds an additional 24 free parameters. The measurement models used in Studies 2-3 are identical, except that they involve only two factors and Measures x_1-x_8.

Appendix C

Items Used to Measure Positive Affect and Negative Affect in Study 3

Adjective Checklist

Active	*Afraid*
Alert	*Hostile*
Determined	*Irritated*
Excited	*Jittery*
Interested	*Nervous*
Proud	*Upset*

Agree-Disagree Response Option (5 points)

I have been feeling very focused and "on task." I've had trouble paying attention. For some reason, I've been feeling sort of nervous. I feel "calm, cool, and collected."

"Describes Me" Response Option (4 points)

I have felt very inspired.
I have had very little interest in things around me. I have felt rather distressed. I have been feeling calm and relaxed.

Unipolar Likert Scale (7 points)

Alert-not alert Enthusiastic-not enthusiastic Distressed-not distressed Scared-not scared