Psychological Bulletin
1999, Vol. 125, No. 5, 611–617

0033-2909/99/$3.00

The Phoenix of Bipolarity: Reply to Watson and Tellegen (1999)

James A. Russell and James M. Carroll
University of British Columbia

In their commentary on our review (J. A. Russell & J. M. Carroll, 1999), D. Watson and A. Tellegen (1999) agreed that when various factors, including activation, are taken into account, the structure of self-reported affect includes a bipolar dimension contrasting pleasant with unpleasant feelings. Agreement on this the central conclusion of our review may surprise readers familiar with the widespread claim that pleasant and unpleasant affect are not bipolar opposites but are largely independent of one another.

Pleasant and unpleasant, relaxed and tense, elated and depressed—such pairs *seem* bipolar. But appearances can be deceptive, and psychometric evidence has challenged their bipolarity, raising fundamental questions that have vexed the psychology of affect for over 40 years now. These questions arise in basic research (how should affect be conceptualized and assessed?) and in applied contexts (are the debilitating effects of negative affect counteracted by, or independent of, increases in positive affect?).

Over the past dozen years, the bipolar view of affect has appeared to be on its deathbed. The question of bipolarity recently returned to center stage in the psychology of affect (Cacioppo, Gardner, & Berntson, 1997, 1999; Feldman Barrett & Russell, 1998; D. P. Green, Goldman, & Salovey, 1993; D. P. Green, Salovey, & Truax, 1999; Larsen & Diener, 1992; Russell & Feldman Barrett, 1999). The result has been enormous progress. Different writers still have different emphases and prefer different terms, but the issues are being clarified, agreement is replacing dispute, new questions are emerging, and the bipolar view is being revived.

Our article (Russell & Carroll, 1999) examined the logic of testing bipolarity in correlational data and used the valence of affect to illustrate the surprisingly complex and potentially confusing issues involved. The question of bipolarity is especially important when raised about valence: Should the ubiquitous pleasant–unpleasant dimension be replaced with two independent dimensions, one representing the intensity of pleasant feelings and the other representing the intensity of unpleasant feelings? Different research traditions have grown up on the basis of opposite answers to this question. Many factors must be considered simultaneously: the actual predictions of a bipolar model, the multidimensional nature of affect, the time frame of assessment, the errors inherent in measurement, the response format used, and the specific affect items involved. Once all these factors are taken into account, a clear answer begins to emerge. The widespread belief that the evidence goes against bipolarity is, we have found, mistaken.

Watson and Tellegen (e.g., 1985) have been influential proponents of independence, and their view of bipolarity and independence is therefore worth examining carefully. Watson and Tellegen (1999) emphasized, and we want to do so as well, that their version of independence does not and never did challenge bipolarity. Indeed, they appear more confident of bipolarity than we are. Watson and Tellegen (1999) wrote that "bipolarity is evident" (p. 601), that "self-rated affect is characterized by a bipolar dimension of pleasant versus unpleasant feeling" (p. 601), that the data demonstrate "the clear existence of a bipolar dimension of happy versus sad affect in self-report data" (p. 604), and that "a bipolar valence dimension already is readily apparent in raw, uncorrected data" (p. 604). They added that "we recently proposed a three-level hierarchical structure. . . . A general bipolar dimension of happy versus unhappy feeling states emerges at the apex of this hierarchy, attesting to its pervasiveness in self-rated affect" (p. 609).

Watson and Tellegen (1999) wrote of a "mutually recognized convergence" (p. 609), and their commentary may well mark an end to the psychometric challenge to bipolarity.[1] To be sure, some details remain controversial. We would tell the history of the field differently, describe the relation between independence and bipolarity differently, use different labels for the dimensions of affect, and offer different suggestions for the future. Still, the controversial parts are mostly peripheral. We cannot help feeling that the remaining disputes have more to do with words than with substance. We appreciate this opportunity to respond to Watson and Tellegen's comments and to clarify several points of misunderstanding. Nonetheless, it would be most unfortunate if quibbles on the sidelines were to obscure the agreement on center stage.

James A. Russell and James M. Carroll, Department of Psychology, University of British Columbia, Vancouver, British Columbia, Canada.

Many messages traveled the Internet between the participants before this exchange was finished. We thank everyone concerned for their effort. We thank James H. Steiger for his advice on certain matters discussed here. Preparation of this article was facilitated by a grant and a fellowship from the Social Sciences and Humanities Research Council of Canada.

Correspondence concerning this article should be addressed to James A. Russell, Department of Psychology, University of British Columbia, 2136 West Mall, Vancouver, British Columbia, Canada V6T 1Z4. Electronic mail may be sent to jrussell@cortex.psych.ubc.ca.

[1] Watson and Tellegen (1999) questioned the bipolarity of certain affective pairs, such as elated versus bored. (Our Figure 1 gave elated as the opposite of depressed.) Russell and Carroll (1999) did not review evidence on this matter, but it should be reviewed. Yik, Russell, and Feldman Barrett (in press) offered evidence in support of the bipolarity of affective dimensions other than valence.

Bipolarity Versus Independence

Two Structural Models

Watson and Tellegen (1999) began by contrasting what at least seem to be "two competing structural models" (p. 601). The first model "assert[s] the importance of a bipolar valence dimension of positive versus negative feeling" (p. 601). There are other dimensions of affect, but, in this model, valence is important and is bipolar. Our review of the available evidence concluded in favor of this first, traditional model (Russell & Carroll, 1999). As we have said, Watson and Tellegen (1999) also endorsed it.

The second, competing structural model "emphasiz[es] the distinctiveness and independence of positive and negative mood states" (Watson & Tellegen, 1999, p. 601). (Their change in wording from *feeling* in the first model to *mood states* in the second is apparently without consequence.) Beginning with Nowlis and Nowlis (1956), researchers including Bradburn (1969) and Costa and McCrae (1980) found psychometric evidence that what had been assumed to be bipolar opposites were in fact correlated only weakly, were distinctive in their relations to other variables, and were therefore better represented as independent of one another. In this currently dominant view, positive and negative affect are two separate dimensions. Our article found little or no psychometric support for this second, competing model.

Now we come to a potentially puzzling part. In their commentary, Watson and Tellegen (1999) seemed to endorse this second model as well as the first. They anticipated that readers might be puzzled by their endorsement of both models: "If proponents of independence can simultaneously argue for the existence of a bipolar valence dimension . . . then the reader may well wonder, 'What exactly is the nature of this ongoing controversy?' " (p. 602). Their explanation was that the opposition between bipolarity and independence is false:[2] "independence and bipolarity are not necessarily opposing or incompatible concepts (unless the bipolar model is unidimensional)" (p. 603). There is a problem with this explanation.

Compatible or Incompatible?

Is bipolarity compatible or incompatible with independence? The answer, of course, depends on just what is said to be independent of what. Two variables that are bipolar opposites are not independent of one another. Two variables that are independent of one another are not bipolar opposites. (The incompatibility of bipolarity and independence in this sense is implicit throughout Watson & Tellegen's [1999] commentary, as when they examine the correlation between two variables to determine whether they are bipolar or independent.) As obvious as this point is, we have two reasons to underscore it.

Our first reason is historical accuracy. When writers such as Nowlis and Nowlis (1956), Bradburn (1969), and Costa and McCrae (1980) made the counterintuitive and controversial claim of the independence of pleasant from unpleasant affect, they meant that what traditionally were thought bipolar opposites are not in fact bipolar but are instead independent of one another. This is the genuine empirical controversy that many scientists have worked on (to cite but a few: Bentler, 1969; Bradburn, 1969; Cacioppo & Berntson, 1994; Costa & McCrae, 1980; Diener & Iran-Nejad, 1986; Diener, Larsen, Levine, & Emmons, 1985; Frijda, 1993; D. P. Green et al., 1993; R. F. Green & Goldfried, 1965; Larsen & Diener, 1992; Meddis, 1972; Parkinson, Totterdell, Briner, & Reynolds, 1996; Russell, 1979; Thayer, 1989). It is the controversy we addressed in our target article. In this sense of independence, bipolarity and independence are incompatible.

Our second reason is that consensus is not necessarily truth, and bipolarity and independence must continue to be examined empirically. This examination requires precise statements. Two variables, x and y, cannot be both bipolar opposites of and independent of one another. Two variables that are bipolar opposites of one another form one single dimension; two variables that are independent of one another form two separate dimensions. Put differently, if x and y are bipolar opposites, the theoretic product–moment correlation between them is -1 (with one type of response format) or $-.467$ (with another). If x and y are independent of one another, the theoretic correlation between them is .00. In this sense of independence, bipolarity and independence are incompatible.

What Watson and Tellegen (1999) seem to have meant is that two variables being bipolar opposites of one another is compatible with two other variables being independent of one another. This, of course, is true, and it brings us to the next question.

Terminology

Just what, according to Watson and Tellegen, is bipolar to what, and just what is independent of what? Watson and Tellegen (1985) wrote that pleasantness is the bipolar opposite of unpleasantness and that positive affect is independent of negative affect. Watson and Clark (1997) wrote that "variations in positive and negative mood are largely independent of one another" (p. 270). Watson and Tellegen (1999) have now clarified their position by reaffirming the bipolarity of the pleasantness–unpleasantness dimension and by renaming the two independent dimensions "positive activation" and "negative activation."

Activation is therefore the key to understanding Watson and Tellegen's answer (and their terminology), as they explained. Activation is the continuum from sleepiness to high arousal thought of as completely orthogonal to the intensity of pleasantness or unpleasantness. By positive affect or positive mood or positive activation, they meant states that were both pleasant and activated (such as enthusiasm). By negative affect or negative mood or negative activation, they meant states that were both unpleasant and activated (such as panic or nervousness).

If we understand them correctly on this, we agree. Indeed, getting to independence from these definitions can be easily done by considering two random variables, x and y, each in standard score form. Assume that the degree of positive valence (x) and the degree of negative valence ($-x$) are bipolar opposites and together independent of the degree of activation (y). If so, the combination

[2] Watson and Tellegen (1999) took the opportunity of this invited commentary to single out D. P. Green, Goldman, and Salovey (1993) as a source of confusion and to criticize them for creating this allegedly false conflict. D. P. Green and his colleagues do not have the opportunity to reply here; so let us add simply that we disagree with Watson and Tellegen's account of history and with their criticisms of D. P. Green et al. See also the exchange between Tellegen, Watson, and Clark (in press) and D. P. Green and Salovey (in press).

of pleasant and activated ($x + y$) is independent of the combination of unpleasant and activated ($-x + y$). When x and y are in standard score form, the strict independence of $x + y$ from $-x + y$ is a mathematical truism.

Watson and Tellegen have not always been understood on this point. For example, the editors of a book in which Watson and Clark (1994) wrote of the independence of positive and negative mood provided this summary: "Watson and Clark . . . put all negative moods into one category and all positive states into another independent mood" (Davidson & Ekman, 1994, p. 94). The independence of their positive mood from their negative mood was achieved precisely by not putting all negative feelings into one category and all positive feelings into another. By their operational definition, positive mood was a subset of positive moods, namely those positive moods combined with high activation. Happiness, serenity, and the like were excluded. Their negative mood was a subset of negative moods, namely those combined with high activation. Depression, melancholy, and the like were excluded.

Getting terminology straight is thus terribly important. Watson and Tellegen's (1999) change in terminology is a much-needed step in dispelling some of the confusion resulting from previous terms. Nevertheless, we suggest one further tiny change to help ensure against confusion. Watson and Tellegen did not specify the bipolar opposites of their new terms, and when they asked of the relation between "positive and negative activation" (see their Table 1 and pps. 604 and 606–607), the unwary reader might have taken that phrase to refer to what are hypothetically bipolar opposites of one another. We therefore suggest the following translation scheme, which also highlights the convergence between their formulation and ours and is based on the very helpful suggestions of Larsen and Diener (1992).

In our terminology (Russell & Carroll, 1999), positive and pleasant are used interchangeably, negative and unpleasant interchangeably. Positive affect, which we also called x, is the bipolar opposite of negative affect, which we also called $-x$. Together, they form one dimension, which we called valence. This is the dimension Watson and Tellegen (1999) called pleasant versus unpleasant feeling. High activation (HighAct), which we also called y, is the bipolar opposite of low activation (LowAct), which we also called $-y$. Together, y and $-y$ form another bipolar dimension, which we called activation. Valence is independent of activation. In developing measuring devices, valence must be assessed in such a way that it is not confounded with high or low activation, and activation must be assessed in such a way that it is not confounded with valence.

In our translation, Watson and Tellegen's positive activation is renamed positive activated affect. It is equivalent to what we called PA/HighAct, meaning positive affect with high activation. It is $x + y$. Its bipolar opposite is negative deactivated affect, or what we called NA/LowAct or $-x - y$.

Watson and Tellegen's negative activation is renamed negative activated affect. It is equivalent to what we called NA/HighAct, meaning negative affect with high activation. It is $-x + y$. Its bipolar opposite is positive deactivated affect, or what we called PA/LowAct or $x - y$.

Watson and Tellegen's positive activated affect is independent[3] of their negative activated affect because $x + y$ is independent of $-x + y$.

Of course, this formulation is a hypothesis. Whether affect is best modeled in these terms, however consensual, remains an empirical question. These new names capture the semantics of affect, but nothing here indicates or assumes that x and y correspond to the biological mechanisms underlying self-reported affect. An important message of Watson and Tellegen's (1985) work on affect must not be obscured: One basic mechanism of affect could be positive activated versus negative deactivated affect, and another could be negative activated versus positive deactivated affect. Underlying biological mechanism is emerging as the question to which the psychology of affect must turn (Cacioppo, Gardner, & Berntson, 1999).

The Independence Model

Return now to the second of two competing structural models that Watson and Tellegen (1999) defined at the beginning of their commentary. That model emphasized the "independence of positive and negative mood states" (p. 601). These words could have two very different meanings: (a) the independence of pleasantness from unpleasantness (when neither is confounded with activation) or (b) the independence of positive activated (vs. negative deactivated) affect from negative activated (vs. positive deactivated) affect.

Independence Model (a) is counterintuitive, is incompatible with the first (bipolar) structural model that Watson and Tellegen (1999) defined, is what Bradburn (1969) and others thought was supported by the evidence, and is what has been the subject of controversy for 40 years. In contrast, Independence Model (b) is compatible with the first (bipolar) model; indeed, it is a corollary of that model. Watson and Tellegen (1999) endorsed (b) and did not endorse (a).

---

[3] We are discussing terminology and not the actual relation between the positive activated affect and negative activated affect axes. The precise theoretical relation between Watson and Tellegen's two major axes is not clear to us. Watson and Tellegen (1999) quoted with approval their own words in which the two axes were said to be "orthogonal" (p. 602). Elsewhere, they described their axes as "45° removed from valence . . . and activation" (p. 603); simple geometry therefore puts their axes 90 degrees apart. They criticized studies that found a significant correlation between their axes on the grounds that "inappropriate content found its way into some of the assessed measures" (p. 605). Thus, in the text, we describe their axes as independent ($r = 0$). On the other hand (see Tellegen et al., in press; Watson, Wiese, Vaidya, & Tellegen, 1999), a more accurate reflection of their current position probably is that the axes are correlated about $-.4$ (which corresponds to an angle of 114 degrees). To create correlated dimensions in our terminological scheme is simple. We created two dimensions by taking unweighted sums ($x + y$ and $-x + y$) of scores in standard form. The correlation between the two dimensions so created can be altered by introducing weights. For example, replace $x + y$ with $.7x + .7y$; replace $-x + y$ with $-.9x + .4y$. The two now correlate approximately $-.4$. On a theoretical level, Watson and Tellegen face a difficult situation because it is not clear how to determine what the theoretical correlation between their axes should be. It is not clear from their perspective what content is appropriate and what inappropriate. From our perspective, a variety of positive activated dimensions could be created and a variety of negative activated dimensions created, and thus a variety of correlations between the two are possible (Russell, Yik, & Steiger, 1999).

Prior to this exchange, we were uncertain whether Watson and Tellegen meant to advocate the independence of what seem to be bipolar opposites, as was done by Nowlis and Nowlis (1956), Bradburn (1969), Costa and McCrae (1980), and others. For example, Watson and Tellegen's (1999) comment on the compatibility of bipolarity and independence suggested to us that perhaps they meant to advocate both the bipolarity and the independence of seeming opposites. In preparing our reply to their commentary, we put this question to Watson and Tellegen, and they replied that they do not now and never did advocate the independence of what seem to be bipolar opposites.[4]

If, when Watson and Tellegen (1985, 1999) wrote of independence, they were not referring to the independence of seeming opposites (even though their former labels implied otherwise), then what did they mean? They were referring to the existence of at least two dimensions in the structure of affect, a point on which there is little or no controversy. Rather than being described as proponents of independence, Watson and Tellegen (1985) might more appropriately be described as proponents of a two-dimensional structure of affect.

Summary

Watson and Tellegen's (1999) commentary, together with other recent articles (Watson, Wiese, Vaidya, & Tellegen, 1999), provided a clear description of their current views. In 1985, they published an influential model of affect anchored by two orthogonal dimensions, one called positive affect, the other negative affect. They included an often-reproduced figure (Watson & Tellegen, 1985, Figure 1, p. 221) that showed a circular ordering of affect variables in a two-dimensional space. They subsequently developed widely used scales to assess the two orthogonal dimensions, scales called the Positive and Negative Affect Schedule (PANAS; Watson, Clark, & Tellegen, 1988). One point of confusion has been that their definition of positive and negative affect was different from the colloquial use of the terms in that Watson and Tellegen were referring to positive or negative affect that was high in arousal. Now, 14 years later, they have modified their two-dimensional structure and their circular (circumplex) ordering of affect variables, acknowledged that the dimensions assessed by their scales are not totally independent, and changed the names they use for various aspects of positive affect and negative affect. Investigators should be careful not to cite them as allies of Bradburn (1969) and others in advocating the independence of what once seemed like bipolar opposites.

When the terminological fog lifts, we can detect no substantive controversy. On the central issue, we are saying the same thing they are, just in different words. Watson and Tellegen (1999) provided no conceptual or empirical challenge to the central substantive conclusion we reached in our review of the evidence, and, indeed, they endorsed it: Pleasant and unpleasant feelings form a bipolar continuum. When they wrote of independence, they meant to assert that affect involves more than one dimension. Once it is understood just what they claim is independent of what, their claim is seen to have nothing to do with bipolarity.

The Influence of Measurement Error

Watson and Tellegen (1999) thought that we overstated the impact of measurement error on observed correlations. Our point was nothing more than what they conceded, namely, that "measurement error produces significant distortions in observed correlations" (p. 603), and "measurement error can be expected to bias observed correlations away from bipolarity" (p. 604).

The practical issue is this: Can measurement error safely be ignored? Watson and Tellegen suggested yes, although they have begun to take measurement error into account in their own recent research (Tellegen, Watson, & Clark, in press). We suggested no, at least for issues surrounding bipolarity. Consider the major dimensions of their structure. Based on uncorrected correlations, which were "invariably low [weak]" (Watson et al., 1988, p. 1065), Watson et al. (1988) presented these dimensions as "orthogonal" (p. 1063), "quasi-independent" (p. 1065), and "independent" (Watson, 1988, p. 1020). However, when measurement error is taken into account, a different conclusion emerges: Data summarized in their own Table 1 show that, far from being independent, these dimensions correlate between $-.43$ and $-.58$ (a highly significant and substantial correlation by most standards in psychology). Either the scales used did not measure the dimensions claimed (a possibility they raised, but see our Footnote 3) or the dimensions are not independent. When the question is the precise magnitude of a correlation, as it is in the debate surrounding bipolarity and independence, it is not safe to ignore measurement error.

Response Format

Response format has emerged as a topic that requires much closer scrutiny, and we agree with much of what Watson and Tellegen (1999) added. We suggested a distinction between strict and nonstrict formats. A strictly bipolar format anchors the response continuum at both ends and succeeds in assigning responses appropriately across a full bipolar continuum. A strictly

[4] Watson and Tellegen also replied that they have always been clear on this point, although they concede that they have been misunderstood. Frankly, we have not found them easy to understand. A number of things puzzle us about their position. First, their position at least appears to have changed in some key ways. For example, Watson and Clark (1997) stated that "oppositely valenced affects tend to be only weakly negatively correlated with one another" (p. 282). In contrast, Watson and Tellegen (1999) stated that "bipolarity is evident even in uncorrected data obtained with unipolar formats" (p. 601), presumably because oppositely valenced affects that are semantic opposites are strongly—rather than weakly—negatively correlated with one another. Second, they aligned themselves with those who did advocate the independence of seeming opposites: Zevon and Tellegen (1982) wrote, "Further evidence in support of distinct dimensions of Positive and Negative Affect has appeared in a number of studies" (p. 121); Zevon and Tellegen then cited Bradburn and Caplovitz (1965) and Costa and McCrae (1980). And third, Watson and Tellegen have not always made it clear precisely what they thought was independent of what. They have sometimes used the terms positive affect and negative affect without making clear that they were to be taken as a combination of valence and activation rather than in their literal meaning; for example, Watson and Clark (1997) defined positive affect as "the extent to which one is experiencing a positive mood" (p. 270) and defined negative affect as "the extent to which one is nonspecifically experiencing a negative or aversive mood." These definitions were followed by the assertion that "variations in positive and negative mood are largely independent of one another" (p. 270).

unipolar format is anchored by an affective state at the high end but by neutrality at the low end; it succeeds in assigning responses appropriately to this unipolar conceptualization. Nonstrict formats appear as their strict counterparts but do not succeed in assigning responses as advertised.

Watson and Tellegen's (1999) analysis using item response theory showed just how problematic a nonstrict format can be: The same nonstrict unipolar response format could be interpreted as bipolar when coupled with the word happy but as unipolar when coupled with the word sad. We have found similar results.

Watson and Tellegen (1999) likely underestimated the impact of different response formats on an observed correlation because they relied on data gathered with the PANAS scales. The problem is that the PANAS does not yield scores of positive or negative affect comparable to what is yielded by scales that include semantic opposites such as happy and sad.[5] Our analysis showed that the influence of strictly unipolar versus strictly bipolar response format on a correlation declines as the true correlation approaches zero. The true correlation between the PANAS Positive Affect scale and the PANAS Negative Affect scale is sufficiently far from -1 (it is approximately $-.4$) that the impact of response format is reduced. For a truly independent pair, the impact disappears altogether. For a truly bipolar pair, the impact is great: A truly bipolar pair can correlate -1.00 with a strictly bipolar format but $-.467$ with a strictly unipolar format. Evidence we reviewed supported this prediction.

Regarding the impact of format on a correlation, there is one point on which we would like to avoid misunderstanding. Watson and Tellegen (1999) characterized our derived correlation of $-.467$ as an "observed coefficient" (p. 605). The value of $-.467$ is predicted to be observed only when certain assumptions are met and when measurement is free of random and systematic error.

Watson and Tellegen (1999) questioned our recommendation that strictly bipolar response formats be used in the routine assessment of affect. For an affective pair whose bipolarity has been established, a bipolar format mirrors the underlying construct and coincides with the respondent's implicit concept. The respondent is given the two extreme anchors and therefore knows how to interpret the question. The researcher knows how to interpret the answer.

Even for bipolar concepts, Watson and Tellegen (1999) preferred a unipolar response format, which, they wrote, "allows the data to 'speak for themselves' " (p. 606). Unfortunately, no format can do that. Watson and Tellegen did not specify whether they meant a strictly or only an ostensibly unipolar format. A strictly unipolar format has its uses, but it does impose a unipolar conceptualization on whatever is assessed. This imposition must be clearly understood and taken into account in interpreting results. A nonstrict format is more difficult to interpret. We found that an ostensibly unipolar format often is ambiguous in that the respondent does not know if the zero point corresponds to affective neutrality or to the bipolar opposite. Different respondents likely interpret the same unipolar format differently. For example, a unipolar response scale labeled *excitement* would be answered one way by a respondent who takes the low end of the scale to mean calm, in another way by a respondent who takes it to mean depressed, and in yet another way by a respondent who takes it to mean an absence of excitement including neutrality. Some formats, although ostensibly unipolar, allow or encourage users to interpret them as bipolar. This ambiguity is a neglected problem in the study of bipolarity.

We made one exception to our recommendation. In a study that seeks to test bipolarity, a bipolar format is inappropriate. Yet such studies are rare, and for several reasons, data gathered in routine assessments of affect ordinarily do not address the question of bipolarity. For example, we showed that in a domain in which bipolarity is assumed, routine factor analysis of data gathered with a strictly unipolar format produces the wrong number of dimensions, with independence appearing as an artifact.

Polychoric Correlation

Watson and Tellegen (1999) have begun using the polychoric correlation. The results are interesting, but caution is in order. Users should make explicit the strong set of assumptions required by polychoric correlations, most importantly that the underlying distribution is bivariate normal and that observed responses are determined by the respondents' "thresholds." The idea is that respondents have thresholds on an assumed underlying normally distributed continuum and that these thresholds convert their responses on that continuum to the observed response scale; the conversion is generally a nonlinear transformation. These assumptions can, and should, be tested (Muthén, 1994). Even so, it must be remembered that the test can only falsify the assumption of underlying bivariate normality, not verify it (Steiger, 1994). Moreover, robustness of the polychoric estimation procedure may vary according to the models being tested as well as the extent to which the data fit those models. Finally, in using a polychoric model in a test of bipolarity, the logic behind that test should be made explicit.

Validity of Affect Ratings for Extended Time Periods

Watson and Tellegen (1999) were puzzled by our remarks on respondents' retrospective ratings of how they felt over a long chunk of time. We did not mean to question the usefulness of such ratings for most purposes, only for testing bipolarity. Thus, Watson and Tellegen argued that the glass we thought half empty is actually half full. Such ratings are neither random numbers nor completely valid. Retrospective ratings are likely less valid than ratings of the current moment. Why? Retrospective ratings are subject to the biases of momentary ratings plus additional biases. Validity of the retrospective ratings presupposes that the rater can validly assess the moment, and further it presupposes that the rater can remember many different moments accurately and then aggregate them appropriately. Affect ratings based on memory are subject to specific known biases (e.g., the reconstructive nature of memory, relative neglect of duration, influence of current mood) that do not exist for affect ratings of the current moment.

Watson and Tellegen (1999) wrote that we "offered no evidence to document [our] claim that general ratings were less valid than

[5] More generally, Watson and Tellegen might want to reevaluate their PANAS scales. The response format used is ambiguous. These scales do not measure the bipolar opposites of pleasant versus unpleasant affect that their title might suggest. These scales do not measure strictly independent dimensions of positive activated and negative activated affect. Other problems with these scales were listed by Larsen and Diener (1992).

other affect measures" (p. 608). (By "general" ratings, they meant retrospective ratings for long, often vaguely specified periods of time.) We cited evidence to support our analysis (Fredrickson & Kahneman, 1993; Schwarz & Clore, 1983) but perhaps could have added references to studies on the reconstructive nature of memory. Alternatively, Watson and Tellegen might have had a different kind of evidence in mind, but we are not sure what that would be.

Watson and Tellegen (1999) reviewed extensive data showing that retrospective affect ratings are not devoid of validity. We agree. Nonetheless, their point is tangential because they did not show (or even claim) that retrospective ratings provide useful information on the questions our article addressed. The data they reviewed neither support nor refute the bipolarity of positive and negative affect. For example, consider the correlation of .58 between the PANAS Negative Affect scale and Neuroticism; we see no bearing of this correlation on the issues we addressed. To be useful to this debate, empirical results must be compared to a prediction correctly derived from a model of bipolarity.

The Circumplex Model

Watson and Tellegen (1999) criticized Russell's (1980) circumplex. Of course, the circumplex is just an approximation. Still, it is a convenient and heuristic one, the best we know of for self-reported affect. Watson and Tellegen acknowledged that certain core features of the circumplex—including bipolarity—are well established. Because bipolarity, and not the circumplex, was the topic of our article, we did not consider the ways that actual data might deviate from a circumplex. The question is clearly important.

Our only quibble with this section of their commentary is that Watson and Tellegen cited only one type of index of fit from one study. Remington, Fabrigar, and Visser (1998) carried out a meta-analysis with three aggregate and 50 individual correlation matrices for affect data. A circumplex typically fit the data well. Indeed, Fabrigar, Visser, and Browne (1997) used affect data to illustrate how well a circumplex can fit actual data.

Back to the Central Issues

These various peripheral disputes should not be allowed to obscure the main points of our analysis, to which Watson and Tellegen agreed or at least raised no objection. We therefore conclude with a counterfactual condition. Suppose that we were to concede every objection they made. This concession would leave unharmed all the major points of our analysis, including these:

1. To test whether any two variables are bipolar opposites, an empirical result must be compared with the precise prediction of an explicit bipolar model for the specific method used to gather the data. This requirement has rarely been met in tests of bipolarity, and therefore much of the available data is of questionable value. (Although Watson and Tellegen did not dispute this principle, they sometimes interpreted observed correlations without making explicit just what model of bipolarity they had in mind. Without such a comparison, the correlation coefficients are just numbers.[6])

2. The observed correlation between positive and negative affect varies with random and systematic errors inherent in measurement, the items used to make the abstract notions of positive and negative affect operational, and the response format. (We would add time frame to this list, but are unsure of Watson and Tellegen's position on this point. We would also add that these influences should not be considered in isolation but cumulatively and interactively.)

3. Two variables that are bipolar opposites can be related to each other in two different ways. If each is defined as the whole bipolar continuum, then their relation is linear (a straight diagonal in a bivariate diagram). If each is defined as a mutually exclusive part of that bipolar continuum, then their relation is nonlinear (an L-shaped pattern in the bivariate diagram).

4. With the exception of Diener and Iran-Nejad (1986), previous analyses of bipolarity of momentary affect suffered from a contradiction, that of requiring both a unipolar response format and a (product–moment) correlation of (approximately) -1. These two requirements cannot be met simultaneously. To achieve a correlation of -1 requires bipolar response formats, yet bipolar response formats are illegitimate in tests of bipolarity. Unipolar formats must be used, but the more strictly unipolar the format, the farther from -1 will be the correlation. When the response format is strictly unipolar, when measures are free from error, when the parent distribution is normal, and when the population mean is zero, then the expected correlation between bipolar opposites is $-.467$.

5. With the exception of Diener, Larsen, Levine, and Emmons (1985), previous writings on the bipolarity of affect extended over time lacked an explicit analysis of the actual predictions of bipolarity. When one person provides repeated momentary affect ratings, the within-subject analysis parallels that for momentary affect. However, when the researcher aggregates those repeated momentary ratings (e.g., taking a frequency or a mean) or when the person is asked to make retrospective ratings about an extended period, bipolarity can yield counterintuitive predictions. For example, any correlation between mean or retrospective mean positive and negative affect (each assessed with a strictly unipolar response format) is consistent with bipolarity.

6. When the available data from studies both of momentary affect and of affect extended over time are compared with the predictions of a thoroughly bipolar model of positive and negative affect, there is no basis whatsoever for a rejection of bipolarity.

---

[6] An analysis of bipolarity and independence was offered by Tellegen, Watson, and Clark (1994), who argued that correlations in the range of -1 to $-.71$ indicate bipolarity and those in the range $-.70$ to .00 indicate independence. Watson and Tellegen (1999) did not follow these guidelines. For example, about correlations ranging from $-.53$ to $-.78$ and having a mean of $-.61$, they wrote, "these data demonstrate that a bipolar valence dimension already is readily apparent in raw, uncorrected data" (p. 604).

References

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

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

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

Cacioppo, J. T., & Berntson, G. G. (1994). Relationship between attitudes and evaluative space: A critical review, with emphasis on the separa-

bility of positive and negative substrates. *Psychological Bulletin, 115,* 401–423.

Cacioppo, J. T., Gardner, W. L., & Berntson, G. G. (1997). Beyond bipolar conceptualizations and measures: The case of attitudes and evaluative space. *Personality and Social Psychology Review, 1,* 3–25.

Cacioppo, J. T., Gardner, W. L., & Berntson, G. G. (1999). The affect system has parallel and integrative processing components: Form follows function. *Journal of Personality and Social Psychology, 76,* 839–855.

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

Davidson, R. J., & Ekman, P. (1994). Afterword: How are emotions distinguished from moods, temperament, and other related affective constructs? In P. Ekman & R. J. Davidson (Eds.), *The nature of emotion: Fundamental questions* (pp. 94–96). New York: Oxford University Press.

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.

Fabrigar, L. R., Visser, P. S., & Browne, M. W. (1997). Conceptual and methodological issues in testing the circumplex structure of data in personality and social psychology. *Personality and Social Psychology Review, 1,* 184–203.

Feldman Barrett, L., & Russell, J. A. (1998). Independence and bipolarity in the structure of current affect. *Journal of Personality and Social Psychology, 74,* 967–984.

Fredrickson, B. L., & Kahneman, D. (1993). Duration neglect in retrospective evaluations of affective episodes. *Journal of Personality and Social Psychology, 65,* 45–55.

Frijda, N. H. (1993). Moods, emotion episodes, and emotions. In M. Lewis & J. M. Haviland (Eds.), *Handbook of emotion* (pp. 381–404). New York: Guilford Press.

Green, D. P., Goldman, S. L., & Salovey, P. (1993). Measurement error masks bipolarity in affect ratings. *Journal of Personality and Social Psychology, 64,* 1029–1041.

Green, D. P., & Salovey, P. (in press). In what sense are positive and negative affect independent? A reply to Tellegen, Watson, and Clark. *Psychological Science.*

Green, D. P., Salovey, P., & Truax, K. M. (1999). Static, dynamic, and causative bipolarity of affect. *Journal of Personality and Social Psychology, 76,* 856–867.

Green, R. F., & Goldfried, M. R. (1965). On the bipolarity of semantic space. *Psychological Monographs, 79* (6, Whole No. 599).

Larsen, R. J., & Diener, E. (1992). Promises and problems with the circumplex model of emotion. In M. S. Clark (Ed.), *Review of personality and social psychology* (Vol. 13, pp. 25–59). Newbury Park, CA: Sage.

Meddis, R. (1972). Bipolar factors in mood adjective checklists. *British Journal of Social and Clinical Psychology, 11,* 178–184.

Muthén, B. O. (1994). Goodness of fit with categorical and other nonnormal variables. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 205–234). Newbury Park, CA: Sage.

Nowlis, V., & Nowlis, H. H. (1956). The description and analysis of mood. *Annals of the New York Academy of Science, 65,* 345–355.

Parkinson, B., Totterdell, P., Briner, R. B., & Reynolds, S. (1996). *Changing moods: The psychology of mood and mood regulation.* London: Longman.

Remington, N. A., Fabrigar, L. R., & Visser, P. S. (1998). *Re-examining the circumplex model of affect.* Unpublished manuscript, University of Maryland Baltimore County.

Russell, J. A. (1979). Affective space is bipolar. *Journal of Personality and Social Psychology, 37,* 345–356.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39,* 1161–1178.

Russell, J. A., & Carroll, J. M. (1999). On the bipolarity of positive and negative affect. *Psychological Bulletin, 125,* 3–30.

Russell, J. A., & Feldman Barrett, L. (1999). Core affect, prototypical emotional episodes, and other things called *emotion:* Dissecting the elephant. *Journal of Personality and Social Psychology, 76,* 805–819.

Russell, J. A., Yik, M. S. M., & Steiger, J. H. (1999). *A 12-point circumplex model of affect.* Manuscript submitted for publication.

Schwarz, N., & Clore, G. L. (1983). Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states. *Journal of Personality and Social Psychology, 45,* 513–523.

Steiger, J. H. (1994). Factor analysis in the 1980's and the 1990's: Some old debates and some new developments. In I. Borg & P. P. Mohler (Eds.), *Trends and perspectives in empirical social research* (pp. 201–223). Berlin, Germany: Walter de Gruyter.

Tellegen, A., Watson, D., & Clark, L. A. (1994, August). Modeling dimensions of mood. In L. A. Feldman (Chair), *Mood: Consensus and controversy.* Symposium conducted at the 102nd Annual Conference of the American Psychological Association, Los Angeles, CA.

Tellegen, A., Watson, D., & Clark, L. A. (in press). On the dimensional and hierarchical structure of affect. *Psychological Science.*

Thayer, R. E. (1989). *The biopsychology of mood and activation.* New York: Oxford University Press.

Watson, D. (1988). Intraindividual and interindividual analyses of positive and negative affect: Their relation to health complaints, perceived stress, and daily activities. *Journal of Personality and Social Psychology, 54,* 1020–1030.

Watson, D., & Clark, L. A. (1994). Emotions, moods, traits, and temperaments: Conceptual distinctions and empirical findings. In P. Ekman & R. J. Davidson (Eds.), *The nature of emotion: Fundamental questions* (pp. 89–93). New York: Oxford University Press.

Watson, D., & Clark, L. A. (1997). The measurement and mismeasurement of mood: Recurrent and emergent issues. *Journal of Personality Assessment, 86,* 267–296.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures 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.

Watson, D., & Tellegen, A. (1999). Issues in the dimensional structure of affect—Effects of descriptors, measurement error, and response formats: Comment on Russell and Carroll (1999). *Psychological Bulletin, 125,* 601–610.

Watson, D., Wiese, D., Vaidya, J., & Tellegen, A. (1999). The two general activation systems of affect: Structural findings, evolutionary considerations, and psychobiological evidence. *Journal of Personality and Social Psychology, 76,* 820–838.

Yik, M. S. M., Russell, J. A., & Feldman Barrett, L. (in press). Structure of self-reported current affect: Integration and beyond. *Journal of Personality and Social Psychology.*

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

Received September 11, 1998
Revision received March 10, 1999
Accepted March 30, 1999 ■