17

Anna Marie Ruef
Robert W. Levenson

Continuous Measurement of Emotion

The Affect Rating Dial

What Is the Affect Rating Dial?

Measurement of an individual's subjective experience of emotion has long been a key component of emotion research, but it presents some unique challenges. Researchers have developed a number of different methods to assess the subjective emotional experiences of study participants, each of which has its strengths and limitations. Self-report measures such as the Positive and Negative Affect Schedule (PANAS; Watson, Clark, & Tellegen, 1988) are well established and easy to complete and provide useful information; however, administration of any written measure necessitates an interruption in the flow of an experiment and does not allow frequent or continuous sampling of affective states. More involved methods, such as interviewing, give a detailed and comprehensive picture of a person's emotions, but they are time-consuming and can provide only a retrospective report of affect. The interactive computer version of the Self-Assessment Manikin (SAM) Scales (Bradley & Lang, 1994) allows online assessment of both emotional valence and arousal levels, but it, too, does not provide a continuous record of affect.

The affect rating dial is an assessment method that was developed to measure self-reported emotion continuously in social interaction. It is unique among subjective measures of affect in allowing the measurement of the time course of an emotional experience and generating continuous data that can be viewed both normatively and ideographically. This chapter outlines the development of the affect rating dial, its strengths and limitations as a measure of subjective emotion, studies in which it has been used, construction of the device, instructions for its use, techniques for evaluating the data it generates, and, finally, possible future uses for the dial.

Development of the Affect Rating Dial

In the early 1980s John Gottman and Robert Levenson were beginning to study the emotional behavior of married couples in the laboratory. They were collecting continuous data on the physiological responses and expressive emotional behavior of their research participants, and they also wished to measure participants' subjective emotional reactions in a similar fashion. They needed to know how couples were feeling as they engaged in two conversations with one another, the first focusing on the events of the day and the second on an area of conflict in their marriage. Ideally, the investigators would obtain a measure of each spouse's subjective emotional state in the moment, as he or she was engaged in the discussion. Also, they wanted to be able to track the rapid changes in emotion they expected to observe (Gottman & Levenson, 1986).

Existing methods for obtaining self-reports of emotion throughout a social interaction required the experimenter to stop the conversation repeatedly to ask participants how they were feeling. Gottman et al.'s talk-table procedure (Gottman et al., 1976; Markman, 1979, 1981) was one such technique: The experimenter would stop the interaction after each communication to obtain spousal ratings of the intent of the message sent and the impact of the message received, each rated on a 5-point Likert scale. However, this method was highly intrusive, and the frequent disruptions resulted in less external validity with regard to real-life marital interactions. The other assessment methods then in use consisted of pre- and postintervention questionnaires and rating scales that did not allow for continuous ratings of emotion. Because the existing methods did not meet their needs, the researchers set out to develop a new procedure for obtaining self-report of affect.

The affect rating dial was a key component of this new "video recall" procedure (Gottman & Levenson, 1985; Levenson & Gottman, 1983). Research participants first came to the laboratory for an interaction session, where they engaged in the events of the day and problem area conversations. A video recording was made of these discussions, and a number of physiological measures were recorded from each spouse. Several days later, participants returned separately to the laboratory to view the videotaped recordings of the conversations, and the same physiological measures were recorded and synchronized with the data obtained previously. While viewing the tapes, participants used the affect rating dial, a type of joystick device, to give a continuous report of their emotions during each marital interaction. The dial traversed a 180° arc over a 9-point scale anchored with the legends *very negative* at 0°, *neutral* at 90°, and *very positive* at 180°. Spouses were instructed to rate how they felt when they were actually in the interaction. They were told to adjust the dial position as often as necessary while viewing, so that it always reflected their emotional state during the discussions.

The investigators hypothesized that participants watching the videotape would at least partially relive the emotions they had experienced at the time of the interactions. If this were true, the ratings obtained with this recall procedure would be as valid and reliable as if spouses' emotions had been assessed during the original session but without the associated difficulties. Five methods were used to assess the validity of the procedure, and all produced supportive data. First, participants' mean affect ratings discriminated high-conflict (problem area) interactions from low-conflict (events of the day) interactions. Second, when mean ratings for the problem-area interaction were correlated with participants' marital satisfaction scores, more dissatisfied couples were found to rate their interaction more negatively. Third, husbands' affect ratings during conversations were found, using time-series analyses, to be coherent with their wives' ratings (see the section on data-analytic approaches later in the chapter). Fourth, participants' ratings were consistent with observers' objective coding of couples' affect (discussed later). Finally, spouses in the recall session appeared to relive the emotions they experienced in the original interaction, as reflected by coherence between interaction session and recall session physiological measures.

Fredrickson and Kahneman (1993) later developed their own variation of the affect rating dial, which they called a *positive-negative affect meter*. This version consisted of a sliding knob attached to a potentiometer that controlled an array of 15 colored lights positioned above the participant's video monitor. A series of 7 green lights to the right of center represented degrees of positive feelings the participant might experience, whereas the 7 red lights to the left of center represented degrees of negative feelings. The yellow light at the center represented neutral feelings and was the only light illuminated when the sliding knob was centered. As the knob was slid to the right, the green lights came on one by one, so that all 7 green lights were illuminated when the sliding knob was set to the extreme right. Negative ratings were indicated similarly by the number of red lights illuminated. A computer recorded the mean position of the sliding knob every second. As with the affect rating dial, participants were asked to adjust this sliding scale as often as necessary so that it always reflected how positive or negative they were feeling. This light display has the advantage of allowing participants to keep their eyes trained on the video monitor while receiving continuous, synchronized feedback on their reported affect level.

Strengths and Limitations of the Method

The affect rating dial as a method of assessment has several positive attributes. First, it allows the investigator to collect online emotion ratings from participants over the time course of an experiment. Its temporal resolution allows participants to indicate rapid changes in their emotions. Second, no forced choices among rating categories are required, because the rating dial provides a continuous measure. A third strength is its ease of use. Once participants are shown how to manipulate the dial, they tend to find it fairly unobtrusive and simple to use. Most are able to make their ratings with only an occasional glance at the device. Fourth, it provides both normative and ideographic data. Finally, it is a robust method that can be validated in a number of different ways when used as part of a video recall procedure (e.g., agreement with objective coders and physiological reliving).

A limitation of the affect rating dial is its use of a single-valence dimension of emotion with two poles (very negative and very positive). Although research suggests that evaluations of positivity and negativity are often combined into a single affective or behavioral (appetitive/aversive) response (Cacioppo & Berntson, 1994; Cacioppo & Gardner, 1999), the rating dial does not allow for those instances in which an individual experiences both positive and negative emotions at the same time. For instance, a research participant who was feeling both mild sadness and mild affection would be forced to choose a dial position that indicates a neutral state, which would not do justice to the complexity of the emotional experience. In this instance, the participant would provide the same rating for that period as would a partici-

pant who chose *neutral* because he or she actually was not feeling any emotion. Another limitation is that the dial provides only a single rating at a time. It may be possible to have participants use multiple dials, but this has not been validated.

In the original studies using the dial (Levenson & Gottman, 1983, 1985), the data it generated were broken down into three categories (neutral, positive, or negative) in order to make comparisons between participants' self-ratings of affect and trained observers' ratings using the Specific Affect Coding System (SPAFF; Gottman & Krokoff, 1989). Dividing the data into categories in this manner leads to a loss of information, but this is not a necessary step in other methods of analysis, such as time-series analysis. By necessity, the rating dial originally was used to measure relived emotions, not emotions as they were actually occurring in an interaction. However, more recent studies show that the dial can be used to provide online reports of affect as it is occurring, such as in film viewing (e.g., Tsai, Levenson, & Carstensen, 2000) without unduly influencing ratings (Mauss et al., 2005). Moreover, although the original studies were limited to ratings of affective valence, a more recent study has used obtained ratings of specific emotions, such as amusement and sadness (Mauss et al., 2005).

Past Uses of the Affect Rating Dial

The affect rating dial has been utilized in a number of different ways in emotion research, for example, in single-participant and dyadic interaction studies; to obtain self-ratings and ratings of a target person's emotions; and to rate emotions online and retrospectively. The following are brief descriptions of research studies that have employed the rating dial and summaries of the results that involve subjective affect ratings obtained in this manner.

Alcohol Studies

The first version of the affect rating dial was used in a series of studies that explored the stress response-dampening effects of alcohol (Levenson, Oyama, & Meek, 1987; Levenson, Sher, Grossman, Newman, & Newlin, 1980; Sher & Levenson, 1982). Instead of reporting their general affective state, participants used the rating dial to provide online continuous ratings of their perceived levels of tension or anxiety throughout the experimental session. This anxiety rating dial was anchored by *extremely calm* (0°) and *extremely tense* (180°), but in other respects it was similar to the later version. After ingesting alcohol or a placebo, participants gave continuous ratings of their tension levels as they were exposed to two stressors, electric shock and a self-disclosing speech. In this context, ratings were reflective of online anxious feelings, not relived emotions. Anxiety dial data were used to demonstrate that alcohol consumption had a dampening effect on subjective tension levels prior to a stressor but not during the stressor.

Dyadic Interaction Studies

With the development of the previously described video recall procedure, the rating dial was adapted to measure the positive-negative dimension of emotion. This version of the dial first appeared in Levenson and Gottman's (1983) study of marital interaction. As described earlier, participants watched a videotape of a conversation with their spouses and continuously rated how they were feeling during that interaction. The dial was used in much the same manner in a series of dyadic interaction studies, both with younger and older married couples (Gottman & Levenson, 1992; Levenson, Carstensen, & Gottman, 1994; Levenson & Gottman, 1985) and dating couples (Tsai & Levenson, 1997).

In the original marital study (Levenson & Gottman, 1983), rating dial data revealed that dissatisfied marital relationships were characterized by less positive and more negative affect and greater reciprocity of negative affect than satisfied relationships; these differences were more pronounced when couples were discussing a highly conflictual topic than one low in conflict. Couples in a 5-year follow-up study (Levenson & Gottman, 1985) demonstrated the greatest decline in marital satisfaction when husbands did *not* reciprocate their wives' negative affect and when wives *did* reciprocate their husbands' negative affect (again, as measured by the rating dial). Older married couples in the long-term marriage study (Levenson et al., 1994) rated their emotions as being more positive and were less physiologically aroused when interacting with one another than were middle-aged couples. Dissatisfied couples in this study reported less positive emotion, more negative emotion, and greater negative-affect reciprocity than did satisfied pairs. Husbands reported feeling more affectively negative the more they were aroused, but wives showed no relation between subjective affect and arousal. Chinese American dating couples who used the rating dial during conversations about an area of conflict in their relationship (Tsai & Levenson, 1997) reported less variable and less positive affect than did European American couples. Rating dial data in each of these studies contributed to a greater understanding of the role of emotion in intimate relationships.

Empathy Studies

In a study of the physiological correlates of empathy, Levenson and Ruef (1992) used the affect rating dial, together with videotape viewing, but in a different manner than in the marital studies. In these studies, participants did not provide ratings of their own emotional states but instead rated how they thought strangers on videotape were feeling. They were asked to watch 15-minute videotaped interactions of married couples, focusing their attention on a designated spouse and using the rating dial to indicate continuously how they thought that spouse was feeling. These ratings were compared with the target spouse's self-ratings, obtained previously using the video recall procedure, to calculate a measure of empathic accuracy. Accuracy scores were correlated with

measures of physiological linkage, or similarity in physiological responding, between the participant during the rating task and the target during the original conversation. Empathic accuracy for negative emotion was associated with a state of shared physiology between the rater and the person being rated, whereas accuracy for positive emotions was associated with a state of low cardiovascular arousal in the rater.

A study of empathy in long-term marriages (Ruef, 2001) required participants to rate both their own emotions and the emotions of their spouses. Spouses in long-term marriages engaged in a 20-minute discussion of a marital problem, then returned separately to the laboratory to view the video recording of their interaction twice. During the first viewing they used the affect rating dial to rate continuously how they had felt during the interaction. Then they were asked to watch the tape a second time and rate how they thought their spouses had felt. Autonomic and somatic physiological responses were monitored continuously during the interaction and rating sessions. As in the previous empathy study with strangers, empathic accuracy was determined by measuring the extent to which the participant's emotion ratings matched those of the spouse; physiological linkage between married partners was determined, as well. For both husbands and wives, a state of low cardiovascular arousal increased the accuracy of detecting both positive and negative emotions in their spouses. Physiological linkage also increased accuracy, but only for wives.

Emotional Responses to Films

Although the affect rating dial has been used primarily in dyadic interaction studies, it has also been utilized successfully in experimental contexts unrelated to social interaction. In a study of duration neglect in global evaluations of affective experiences, Fredrickson and Kahneman (1993) used their version of the rating dial to measure affective escalation and satiation over the course of viewing emotional film clips. They found that the duration of the viewed emotional episodes had little effect on participants' retrospective evaluations of their feelings. Later, Fredrickson and colleagues (Fredrickson & Levenson, 1998; Fredrickson, Mancuso, Branigan, & Tugade, 2000) used the rating dial to explore the "undoing" effect of positive emotions on cardiovascular arousal. The dial was used to measure participants' subjective emotional states as they watched affectively positive and negative film clips. Tsai et al. (2000) utilized the dial in a similar fashion in their study of older and younger Chinese Americans' and European Americans' responses to emotional films. Rating dial data revealed no age or cultural differences in subjective responding to the films. These uses are similar to the original use of the anxiety rating dial in the previously mentioned alcohol studies (Levenson et al., 1980; Levenson et al., 1987; Sher & Levenson, 1982). They represent a return to the online measurement of a participant's subjective feelings when presented with an emotion-inducing stimulus. Concerns about whether providing the ratings influenced the emotional reactions were addressed in a recent study (Mauss et al., 2005), which found no meaningful differences between online ratings and video recall ratings.

Implementation of the Affect Rating Dial

Construction of the Device

The affect rating dial was originally constructed in a 7.5 x 5 x 3.25-inch (19 x 12.5 x 8 cm) metal box. The box had a faceplate with the rating scale and a large knob. All of the electronics were housed inside the box. The box attached to the arm of the participant's chair, so that the participant's hand could rest comfortably on the dial and manipulate it easily. The box could be moved to either arm of the chair, depending on the handedness of the participant.

Faceplate

The faceplate consisted of an 180° arc drawn in black against a white background. Running along the arc were nine tick marks, one every 20°, labeled "1" to "9." The arc was anchored with the legends *very negative* at 0°, *neutral* at 90°, and *very positive* at 180° (see Figure 17.1). This faceplate was covered by a rectangle of clear Plexiglas or plastic.

Dial

The dial consisted of a large round plastic knob (2.4 inches, or 6 cm, in diameter) with a clear Plexiglas pointer attached. As the knob rotated clockwise, it caused the pointer to traverse the 180° arc scribed on the faceplate.

Mounting to Chair

The rating dial box was mounted atop a wooden "arm" that slid into a metal sleeve attached to the arm of the participant's chair. (Sleeves were attached to both arms of the chair to accommodate right- and left-handed participants.) When the wooden arm was slid into place, the box rested against the

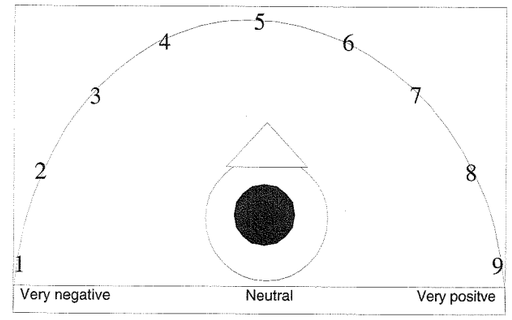

Figure 17.1. The affect rating dial faceplate.

end of the arm of the chair, so the participant's hand could rest easily on top of it to manipulate the dial (see Figure 17.2). Other methods may be used to mount the box and the dial and pointer, keeping in mind participant comfort, need for easy accessibility, and adjustment for handedness.

Electronics

The electronics for the dial consisted of 5-volt DC power applied across a 100K linear potentiometer. The dial pointer was attached to this potentiometer, with the wiper contact and ground leading to a 12-bit analog-to-digital converter on the laboratory computer. This voltage-dividing circuit provided a signal to the computer system that was proportional to the dial position. The dial was calibrated by determining what the voltage readings were when the knob was turned completely to the left and completely to the right. The linear potentiometer allowed the researcher to interpolate the readings by calculating a simple ratio.

Interface With Computer

The laboratory computer used in conjunction with the affect rating dial sampled the dial position at a rate of 30 times per second and then computed the average dial position every second. Because this same computer monitored the physiological measures, whenever rating dial and physiological measures were obtained at the same time, synchronization was ensured. Initially, to synchronize these data with the videotapes of participants' behavior, a tone was recorded on an unused audio track at the precise start of each experimental trial, and an on-screen timer was started. This tone could then be detected by the computer during the video recall session to start the timing of the rating dial and physiological data. Later, this system was replaced by one that recorded a unique number in the vertical retrace interval of every frame of the video. Whereas the older system allowed synchronization to be checked only at one point in time per trial (i.e., at the start of the trial), the new system allowed this throughout the trial.

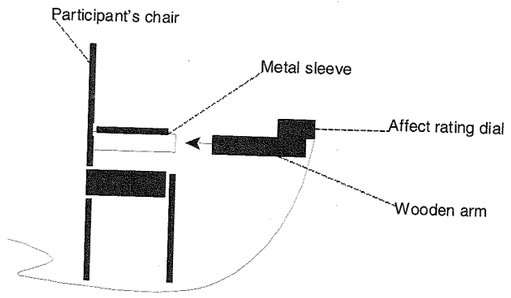

Figure 17.2. Mounting the affect rating dial. A metal sleeve is attached to the underside of the arm of the participant's chair. The dial is mounted atop a wooden arm, which slides into the sleeve.

The original rating dial was an electromechanical device. Recently, Robert Levenson's laboratory has developed a version that is implemented entirely on a computer. The participant sits at a computer screen on which the videotape being rated is played back in a large window. Below the window a horizontal "scroll bar," similar to those used in browser windows, with the numerical 9-point scale and anchor labels provides visual feedback. Participants move the bar with the computer's mouse to indicate changes in their emotional state. This method requires no external electronics at all, and we have found that participants seem quite comfortable with this approach.

Instructions for Participants

The following are sets of specific instructions for the use of the affect rating dial. Typically, the experimenter or research assistant would deliver one of these explanations in person to the participant, as a part of the general experimental instructions. These are meant to be a guide for those who plan to use the rating dial and will need to be altered depending on the experimental context.

For Dyadic Interaction Studies

I'm going to ask you to use this rating dial with this hand [indicate dominant hand] during the experiment. Please use it to give a continuous report of how you were *feeling* during the discussion with your spouse. You can choose between "extremely positive" [pointing the dial to each position as it is mentioned], "extremely negative," neither positive nor negative [pointing to "neutral"], or *any point in between;* it's a continuous scale. It's like keeping a radio station tuned in or using a steering wheel to stay on course. Just keep your hand on the dial and keep it adjusted to show how you were feeling *at the time of the interaction, not* how you feel about watching yourself on videotape. Do you have any questions about what you are supposed to do? I need to go next door now and make some more adjustments. While I'm gone, I'd like you to practice using the rating dial to get used to it. That will also help me to see if it's registering properly. I'll let you know when I'm ready to start the videotape. [Prior to starting the first video segment, the experimenter lets the participant know that the tape will be starting and reminds him or her of the task.] Please adjust the dial as often as you need to so that it indicates how you were feeling during the entire conversation.

For Empathy Studies

I'm going to ask you to use this rating dial with this hand [indicate dominant hand] during the experiment. Please use it to give a continuous report of how you think the person you are watching is *feeling* during the 20-minute tape. You can choose between "extremely positive" [pointing the dial to each position as it is mentioned], "extremely negative," neither positive nor negative [pointing to "neutral"], or *any point in between;* it's a continuous scale. It's like keeping a radio station tuned in or using a steering wheel to stay on course. Just keep your hand on the dial and keep it adjusted to show how you think the person is feeling. *Don't* base your ratings on how *you* would feel if you were in the interaction. You should pay careful attention to the target person and try to determine how *he or she* was feeling. Do you have any questions about what you are supposed to do? I need to go next door now and make some more adjustments. While I'm gone, I'd like you to practice using the rating dial to get used to it. That will also help me to see if it's registering properly. I'll let you know when I'm ready to start the videotape. [Prior to starting the first video segment, the experimenter lets the participant know that the tape will be starting, and reminds him or her of the task.] Please concentrate on the [man/woman] in this segment and indicate with the dial how you think they are feeling during the entire 20 minutes.

For Film Viewing Studies

(The instructions for these types of studies are similar to those used in the dyadic interaction studies, but with the substitution of present-tense language, e.g., ". . . a continuous report of how you *are feeling* as you view the film clip." The following are some additional instructions used by Fredrickson and Levenson in their 1998 study.)

Negative and positive can mean a lot of different things. In this context, we'd like you to consider "Positive" as referring to any positive emotions such as amusement, contentment, happiness, or calmness and "Negative" as referring to any negative emotions, such as sadness, anger, disgust, frustration, irritation, fear, or contempt.

Data Analytic Approaches

Affect rating dial data can be analyzed in a variety of ways, depending on the nature of the experimental context in which they were gathered and the types of research questions posed. The following section briefly describes each analytic approach and gives examples of its application in past studies that utilized the rating dial.

Mean Affect Ratings

The simplest way to analyze affect rating dial data is to calculate a participant's mean emotional response over an entire interaction or stimulus exposure. In the original study that made use of the dial (Levenson & Gottman, 1983), mean rating dial scores were computed for each 15-minute marital discussion. These means were used to determine whether the study's conversation topics elicited different levels of negative affect. As the investigators had hoped, *t* tests revealed that the couples' average rating for the conflict discussion segment was significantly more negative than the average rating for the events-of-the-day segment. Simple means were also used to compare the husbands' and wives' overall affect ratings within each discussion.

In the long-term marriage study (Levenson et al., 1994), preconversation rating dial means were subtracted from conversation means to give a change score. This score was analyzed in univariate analyses of variance (ANOVAs) with age (middle-aged couples vs. older couples) and marital satisfaction (satisfied vs. dissatisfied) and correlated with spouses' mean physiological levels.

In a similar fashion, Tsai et al. (2000) used overall mean rating dial scores in their study of age and ethnic differences in responding to emotional films. They calculated mean levels of rating dial response for the one-minute prefilm baseline and means of participants' responses during the film viewing period, then subtracted the baseline means from the film means. The resultant change scores were used in later analyses to control for age differences in responsiveness during the baseline period. The investigators conducted an ANOVA for rating dial change scores to examine the effects of culture, age, and type of film on participants' subjective reports of emotion.

Fredrickson et al. (Fredrickson & Levenson, 1998; Fredrickson et al., 2000) used rating dial data to determine whether the film stimuli used in their studies induced or altered subjective emotional states the way the experimenters intended. In the first pair of studies (Fredrickson & Levenson, 1998), they calculated each participant's mean rating dial response over a prefilm baseline, then over the entire period of each stimulus film. For those participants whose means represented significant changes from baseline (determined by within-subject *t* tests), they also determined the peak response during the fear film and the latency to achieve it. An omnibus ANOVA, followed by planned comparisons, confirmed that the affective responses to several secondary films differed from each other and from the initial negative film clip. In a second pair of studies (Fredrickson et al., 2000), the researchers calculated change scores by subtracting baseline period means from film period means and then ran within-subject *t* tests to examine whether these change scores were significantly different from prefilm resting baselines. An ANOVA was used to explore sex and ethnicity differences in rating dial responses.

Although the use of mean affect ratings offers the advantage of simplicity, it does not take advantage of a major strength of rating dial data, the continuous measurement of subjective emotion over time. The following techniques make full use of data obtained from this method.

Amount-of-Affect Scores

In addition to simple means, investigators in the marital interaction studies (Levenson & Gottman, 1983; Levenson et al., 1994) calculated variables that they called "amount of affect" scores. These scores reflected the amount of negative and positive affect for each participant, taking into account the participant's own range of affect ratings. Analyzing rating dial data in this way appears to offer unique information not pro-

vided by mean dial position alone: In the 1994 study, correlations between these two types of variables suggested a sizeable amount of unshared variance (approximately 30–80%).

For the first step in computing amount-of-affect scores, means and standard deviations were calculated for each spouse's rating dial data for baseline segments and for each of the 15-minute discussions. Next, spouses' raw-score dial ratings for each 15-minute marital conversation were averaged into 90 10-s periods. *Z* scores were then computed for each of these 90 periods, using the mean and standard deviation for the 5-min baseline preceding that segment (Levenson & Gottman, 1983) or the mean and standard deviation of the 90 interaction periods (Levenson et al., 1994).

Once these basic computations were made, the next step involved classifying the raw-score average for each 10-s period as positive, negative, or neutral. To be coded positive, a raw-score average had to be greater than or equal to 6.0 (referenced to the 1–9 affect rating scale) and the *z* score had to be greater than or equal to 0.5. For a given period, a positive code meant that the pointer was actually on the positive portion of the dial (the raw-score criterion) *and* was positive relative to the participant's range of ratings (the *z*-score criterion). To receive a negative classification, the raw-score average had to be less than or equal to 4.0 and the *z* score had to be less than or equal to –0.5.

For the final step, two amount-of-affect variables were calculated for each spouse: the number of positive periods and the number of negative periods. Correlating these scores with marital satisfaction scores enabled the researchers to observe a pattern of less positive and more negative affect in dissatisfied marriages (Levenson & Gottman, 1983). Amount-of-affect scores were analyzed in a similar manner to rating dial change scores in the 1994 marital study (i.e., in ANOVAs with age and marital satisfaction and correlated with participants' mean physiological levels).

Rating dial *z* scores were used later to assess the validity of the rating dial/video recall procedure (Gottman & Levenson, 1985). For this analysis, dial ratings for 10-s periods were classified as positive or negative based on the *z*-score criterion alone, rather than on a combination of raw and *z* scores. Observers used the SPAFF coding system to categorize couples' speech units as positive, negative, or neutral, based on a speaker's verbal content, voice tone, context, facial expression, gestures, and body movement. Coders' data were converted to the proportion of negative affect in each 10-s period (i.e., the number of each spouse's speech units coded as negative in the period divided by the total number of that spouse's speech units in the period) and the proportion of positive affect in each period. Investigators compared the agreement between spouses' affect dial ratings for each 10-s period and observer's objective coding of the speech units for both spouses within that period. For example, if a 10-s period were rated negatively by a husband, they predicted that a greater proportion of his speech units and his wife's speech units would be coded negatively by the observer, compared with 10-s periods rated positively by the husband. Repeated measures ANOVAs were used to test these hypotheses, and the relations between subjective rating dial data and objective coding were found to be significant in all cases (i.e., for each spouse and for positive and negative affect).

Lag Sequential Analysis

Lag sequential analysis is a more complex method of analyzing rating dial data that was employed in Levenson and Gottman's first (1983) marital interaction study and in the empathy study (Levenson & Ruef, 1992) described earlier. In the marital work (Levenson & Gottman, 1983) this method was used to study "affect reciprocity," or patterned exchanges of affect between spouses. The first steps in this analysis were outlined in the previous section on amount-of-affect scores; they involved classifying 10-s periods as positive or negative. In the next step, affect-reciprocity scores were calculated for positive and negative affect at lag 0 (i.e., both spouses gave the same rating in the same 10-s period) and lag 1 (i.e., one spouse's rating in a given 10-s period was matched by the other spouse's rating in the following period). A match between spouses for any given 10-s period required both to have rated the period positive or negative; neutral ratings were not counted as matching either positive or negative ratings.

Each affect-reciprocity *z* score was computed by subtracting the unconditional probability from the conditional probability and dividing by an estimate of the standard error. For example, the formula for the wife reciprocating the husband's positive affect at lag 0 was

$$\frac{(HWPOS/HPOS) - (HPOS/90)}{SQRT\{(HPOS/90) \times (1-[HPOS/90])/HWPOS\}},$$

where HWPOS = number of periods in which *both* the husband's and the wife's affect ratings were coded positive, HPOS = number of periods in which the husband's affect rating was coded positive, SQRT = square root, and 90 = total number of 10-s periods in the 15-min interaction. The formula for the husband reciprocating the wife's positive affect at lag 0 was

$$\frac{(WHPOS/WPOS) - (WPOS/90)}{SQRT\{(WPOS/90) \times (1-[WPOS/90])/WHPOS\}},$$

At lag of one period, the formula for the wife reciprocating the husband's positive affect was

$$\frac{(HWPOS1/HPOS) - (HPOS/90)}{SQRT\{(HPOS/90) \times (1-[HPOS/90])/HWPOS1\}},$$

where HWPOS1 = number of periods in which the husband's affect rating was coded positive and the wife's affect rating was coded positive in the *following* period. The formula for the husband reciprocating the wife's positive affect at lag 1 was

$$\frac{(WHPOS1/WPOS) - (WPOS/90)}{SQRT\{(WPOS/90) \times (1-[WPOS/90])/WHPOS1\}}.$$

Similar formulas were used to compute the *z* scores for negative affect reciprocity. These formulas corrected for the total number of periods that met the positive and negative coding criteria, and the resultant scores can be seen as indications of the gain in prediction of one spouse's affect by knowledge of the partner's affect (see Allison & Liker, 1982, for a discussion of the advantages of alternative algorithms). Affect-reciprocity *z* scores were correlated with couples' marital satisfaction scores to examine the affect patterns in distressed and nondistressed marriages.

Affect dial data was utilized similarly in the study of the physiological substrate of empathy (Levenson & Ruef, 1992). The researchers wished to measure participants' accuracy in determining the feelings of targets they viewed on videotape. Instead of computing affect-reciprocity scores, they used lag sequential analysis to compute rating accuracy scores for participants at lag 0 and lag 1. In addition, they examined lag minus 1 (participant matches target's rating for the following 10-s period) to determine whether participants' accuracy was based on actually viewing the targets' behavior, as opposed to a chance relation between the way participants and targets used the rating dial.

Formulas for deriving rating accuracy scores were adapted from the preceding affect-reciprocity formulas, with the substitution of the target's data for the spouse's. For instance, the formula for a participant's accuracy of rating positive affect at lag 1 was

$$\frac{(TSPOS1/TPOS) - (TPOS/90)}{SQRT\{(TPOS/90) \times (1-[TPOS/90])/TSPOS1\}},$$

where TSPOS1 = number of times in which the target's affect rating was positive in a given period and the participant's affect rating was positive in the following period; TPOS = number of periods in which the target's affect rating was positive; and SQRT = square root. Similar formulas were used to calculate rating accuracy for negative affect at the various lags. Rating accuracy scores were correlated with an overall index of physiological linkage, linkage *z* scores for individual physiological variables, mean physiological levels and variabilities, and scores on traditional empathy scales.

The researcher who uses the preceding formulas to assess affect reciprocity or rating accuracy should take care to examine the distribution of *z* scores computed from them. In the long-term marriage study (Levenson et al., 1994), these scores showed greater departure from normality than did the sequential probabilities themselves, so the investigators made the decision to use the latter for affect-reciprocity scores. For instance, positive affect reciprocity at lag 0 would be the number of positive periods for which the other spouse also rated the period as positive (i.e., HWPOS or WHPOS). This method has the advantage of simplicity and ease of interpretation, but it does not correct for total number of periods rated positive or negative.

Alternate Rating Accuracy Indices

Percentage Index

In their empathy study, Levenson and Ruef (1992) also used an index based on the percentage of matching ratings to estimate participants' mean overall level of rating accuracy. This was done by calculating the percentage of 10-second periods rated positive by the target person that were also rated positive by the participant in the same rating period. Similarly, a percentage was computed for participants' matching targets' ratings from the *previous* rating period (lag 1). Eight such percentages were derived (for positive or negative affect in the first or second conversation rated and at lag 0 or lag 1); each served as a simple index of rating accuracy. The mean level of this accuracy index across participants ranged from 28 to 43%, with the majority above the chance level (33%), whereas the performance of individual participants ranged from 0 to 100% accuracy. Participants' rating accuracy indices were correlated with measures of physiological linkage between participant and target and with measures of participants' cardiovascular arousal.

Mean Square Difference Index

A different rating accuracy index was used to determine which videotaped marital conversations to use as stimuli in the empathy study. The researchers wanted to avoid target spouses whose self-ratings were highly idiosyncratic, because others probably would not be able to rate them at all. They had a group of participants rate a number of conversations using the affect rating dial and then determined agreement by examining similarity between targets' and participants' ratings in two different ways. They took into account both similarity in mean affect ratings over the 15-minute conversation and similarity in second-to-second variation, using a simple index based on the mean square differences.

The researchers decided to compare ratings not only in the same time frame but also lagged by increments of 1 second. A maximum lag of 25 seconds in either direction was chosen to allow for delays in detecting the target's affect, as well as anticipation of affect. The difference between the participant's ratings and the target's ratings were calculated for each lag time, and these differences were averaged, producing a mean raw difference score for each lag. The best of *these* was called the best raw difference score. The same procedure was used with *z* scores to derive the best *z* difference score. The average closest raw difference score was obtained by comparing the participant's rating for each second with

all target ratings in the corresponding 50–s window and finding the closest match; these "least differences" were then averaged. The average closest z difference score was derived in the same manner, using z scores. The sum of the preceding five variables (best raw difference + best z difference + average closest raw + average closest z + mean difference) was divided by the standard deviation of the target person's ratings to produce an index of rating accuracy. This index allowed the researchers to select for use in the empathy study those four target spouses whose self-ratings best agreed with those of the participants.

Spectral Time-Series Analysis

Another approach to making sense of affect rating dial data is spectral time-series analysis. Unlike time-series regression based on generalized least squares, spectral time-series analysis allows the researcher to consider all lags simultaneously. It produces a coherence statistic for each frequency in the overtone series; together, these measure the degree of linear association between two time series. Gottman and Levenson (1985) used this technique with marital data to support the validity of their video recall procedure. They examined the relationship between a husband's and wife's affect ratings, and in a second analysis they compared a spouse's physiological responses in the interaction session with his or her responses in the recall session. For each analysis, the two compared series were considered related if the coherence was significant only in the frequency range of maximum variance for both series.

As a first step in this analysis, affect rating dial data were processed as previously described (see the earlier section on amount-of-affect scores), then each spouse's 10-s raw-score data were transformed into a 10-s z-score time series using the mean and standard deviation for that spouse during that interaction segment. Next, spectral density functions were derived for each time series, and the frequency range that contained the maximum variance for both series was identified. The coherence spectrum within this frequency range of maximum overlap was examined to determine whether it was significant. If it was, the peak value of the coherence within this band was recorded; if not, zero coherence was recorded. The investigators then performed a binomial sign test across couples: the z score for this test determined whether the number of couples showing significant coherence for a given measure was significantly greater than chance. They made the conservative assumption that significant coherences would be found in half of the couples by chance. Finally, they reported the averages of the maximum coherence across couples (analogous to a Pearson r^2). Results of these spectral time-series analyses demonstrated coherence between the husband's and wife's self-reports of affect for a given interaction; physiological data also reflected coherence between physiological behavior in interaction and recall sessions.

Although spectral analysis is appropriate for data collected in many experimental contexts, there are certain circumstances for which it is not adequate. For example, this analysis should not be used to test for dominance (one participant's behavior influencing the responses of another) and bidirectionality (each participant's behavior exerting influence on the responses of the other) because it does not control for autocorrelation (Gottman & Ringland, 1981). In cases in which the researcher wishes to test for cross-correlation but suspects that autocorrelation may be a problem, bivariate time-series analysis is a better choice.

Bivariate Time-Series Analysis

Bivariate time-series analysis attempts to account for as much of the variance in a given series (e.g., husband's heart rate) as is possible by knowledge of its past (i.e., the autocorrelation) and then determines how much additional variance can be accounted for by adding knowledge of the past of the other series (e.g., wife's heart rate). In other words, the past of one series is used to predict the residual from the autoregression of the other series (Gottman & Ringland, 1981). In the marital and empathy studies previously described (Gottman & Levenson, 1985; Levenson & Gottman, 1983; Levenson & Ruef, 1992; Ruef, 2001), bivariate time-series analysis was used with physiological data to control for cyclicity in each participant's bodily responses. Also, this form of analysis was considered appropriate because physiological data gathered at different times from the same participant or from two participants in an interaction are not independent observations (Levenson & Ruef, 1992). Bivariate time-series analysis was not typically used with rating dial data in these studies, for two reasons: first, cyclicity is not as characteristic of rating dial output; and second, neither the partner nor the target were in the room to influence the participant's ratings during the video recall or rating session, when rating dial data were collected. However, if dyadic study participants in recall sessions are indeed reliving emotions experienced at the time of the interaction with their partners, one could make the case that partner ratings are not independent of one another and do merit a bivariate time-series approach.

This analysis uses the 10-s-period z-score averages (see the previous section on amount-of-affect scores) to produce two log-likelihood statistics for each variable obtained from the participant and the partner or target. These log-likelihood statistics, denoted Q, have approximately chi-square distributions. For each variable, the first of these statistics represents the extent to which a participant's pattern of response accounts for variance in the partner's pattern of response beyond the variance accounted for by the partner's autocorrelation. The second represents the extent to which the partner's pattern of response accounts for variance in the participant's pattern of response beyond the variance accounted for by the participant's autocorrelation. The Q values can be tested for significance, or they can be converted to z scores to allow comparison across experimental measures. For more detailed information on bivariate time-series analysis, the reader is referred to Gottman and Ringland (1981).

Future Uses of the Affect Rating Dial

Research participants have used the affect rating dial to rate their own emotions and the emotions of others in a variety of experimental contexts, as described previously. However, many other applications are possible. For instance, the dial has been utilized to study emotion in marital interaction, but it would work equally well in studies of other dyads, such as pairs of friends, therapist and client, or mother and child. Use of the dial has been limited to adult participants, but its ease of use makes it appropriate for studies that focus on the affective experiences of children or adolescents. Participants have used the dial to rate their own emotions and those of others; coders could use it in the same manner to rate the affect of participants.

Although the affect rating dial can provide useful data on its own, it is used most effectively in combination with other measures of emotional responding to allow exploration of the relationship between subjective affect and observed emotional behavior or physiological changes. Data from systems that quantify emotional behavior, such as the SPAFF or the Facial Action Coding System (FACS; Ekman & Friesen, 1978), can be paired with rating dial data for this purpose. Other methodologies that may be used with the dial are those that measure electrical activity in the brain, such as electrophysiological measures of prefrontal asymmetry or functional magnetic resonance imaging (fMRI). For example, Larson, Sutton, and Davidson (1998) have studied the relationship of differential brain activation to the time course of recovery from an emotional challenge, as measured by startle magnitude. In future studies, rating dial data collection could be added to these measures to study the time course of recovery from subjective emotional arousal, and then the patterns of recovery in different response systems could be compared for coherence. A continuous measure of subjective affect would seem a useful addition to this and other studies of "affective chronometry" (Davidson, 2000).

In addition to dyadic interaction and emotion elicitation using film clips, the rating dial could be used to evaluate the effects of other emotion elicitors. Participants could use the rating dial to give continuous ratings of affect when presented with emotionally evocative still images or sounds, such as the International Affective Picture Series (IAPS; Lang, Bradley, & Cuthbert, 1995) or the International Affective Digitized Sounds (IADS). Obtaining continuous ratings of subjective emotion also would be useful when participants are asked to vividly imagine an emotional scene, such as in personalized script-driven imagery procedures (Lang, Levin, Miller, & Kozak, 1983; Pitman & Orr, 1986; Pitman, Orr, Forgue, de Jong, & Claiborn, 1987). Another possible application of the rating dial is in research in which participants pose emotional facial expressions and body postures. Use of the dial under such conditions may prove somewhat challenging, depending on the difficulty of the pose; in some cases the participant might not be able to give ratings without interfering with the manipulation. In contrast, the affect rating dial seems especially suited to studies that use music as an emotion elicitor, because the emotional content often varies over time in musical selections and because the aural stimuli and the visual ratings would not interfere with one another. If not too intrusive, the rating dial could easily be added to many studies, to give a more immediate and comprehensive rating of participant's subjective emotional states than could be gathered from traditional retrospective self-report measures.

Thus far, the discussion of alternate uses of the rating dial has focused on ratings of general affect, but with a few simple modifications the dial has been used to measure the level of a specific emotion such as anxiety, amusement, or sadness (Levenson et al., 1980; Mauss et al., 2005). In these uses the dial scale reflects the intensity of the single emotion rather than a bipolar measure of positive and negative affect. An example of such an application would be using the dial to rate subjective anxiety during exposure to feared stimuli in individuals with anxiety disorders, such as social phobia or posttraumatic stress disorder.

Rating more than one discrete emotion presents a different kind of challenge. As we have seen, an important aspect of rating dial application is the time frame of data collection: online versus retrospective (relived emotions). In studies in which a participant is interacting with a partner or engaged in any activity that requires a great deal of concentration, then rating dial data are best gathered retrospectively, using the video recall procedure. When the experimental task is more passive, not requiring a great deal of physical activity, speech, or interaction with others, then online emotion ratings are more feasible. Because the investigator is asking participants to rate their emotional state at the same time as they are attending to a stimulus, there are limits on how much information they can be expected to process at one time. With video recall, participants could watch the videotape of their interaction more than once, giving ratings of different discrete emotions on each pass. With online ratings, the computerized, mouse-driven affect rating dial could be programmed to toggle regularly from the measurement of one emotion to another, perhaps changing the color of the pointer (e.g., red to blue) to indicate the change. For example, the investigator could collect 1 minute of data for anger, then 1 minute for another emotion, such as sadness. However, with this kind of sampling one would lose the continuity of the data stream. There is always a balance between what a participant can reasonably attend to without becoming distracted from the experimental task at hand and the level and amount of information the experimenter wishes to gather. If the rating task is too complex, one risks distraction and intrusiveness; if too simple, one is left with

a measure that offers little emotional information and no continuity of data.

The rating dial also could be modified to measure aspects of emotion other than valence, such as perceived level of physiological arousal. Dimensions of metamood (Mayer & Stevens, 1994; Salovey, Mayer, Goldman, Turvey, & Palfai, 1995), such as acceptance and comfort with one's emotional state, could be tracked over the course of an experiment and related to subjective affect ratings or physiological measures.

Conclusion

In their article demonstrating the validity of the affect rating dial and the video recall procedure, Gottman and Levenson (1985) concluded:

> We expect that the procedure will also be valid for other kinds of dyadic interaction besides marital interaction as long as the interaction produces a reasonable range of emotional responding. It is our hope that the strength of these findings will provide other researchers with sufficient confidence in these procedures to be able to adopt them when a continuous self-report of affect is needed. For those experimental paradigms that differ substantially from ours or in instances where the amplitude of emotional responding is suspect, we have provided two validational techniques—agreement with objective coders and physiological reliving—that could be utilized to validate our affective self-report procedures in other experimental contexts. (p. 159)

Since the time this statement was written the affect rating dial has continued to play a part in studies of interaction, but it also has been utilized in experimental contexts that the authors may not have anticipated. These applications were outlined in this chapter, along with possible future implementations of the measure.

The years since the affect rating dial was developed have seen a great expansion in emotion research, including techniques for emotion assessment. We hope that the information contained in this chapter will be helpful to those investigators looking for ways to continuously measure subjective emotion and will stimulate new creative applications and adaptations of the rating dial.

References

Allison, P. D., & Liker, J. K. (1982). Analyzing sequential categorical data on dyadic interaction: A comment on Gottman. *Psychological Bulletin, 91,* 393–403.

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry, 25,* 49–59.

Cacioppo, J. T., & Berntson, G. G. (1994). Relationship between attitudes and evaluative space: A critical review, with emphasis on the separability of positive and negative substrates. *Psychological Bulletin, 115,* 401–423.

Cacioppo, J. T., & Gardner, W. L. (1999). Emotion. *Annual Review of Psychology, 50,* 191–214.

Davidson, R. J. (2000). Affective style, psychopathology, and resilience: Brain mechanisms and plasticity. *American Psychologist, 55*(11), 1196–1214.

Ekman, P., & Friesen, W. V. (1978). *The Facial Action Coding System.* Palo Alto, CA: Consulting Psychologists Press.

Fredrickson, B. L., & Kahneman, D. (1993). Duration neglect in retrospective evaluations of affective episodes. *Journal of Personality and Social Psychology, 65*(1), 45–55.

Fredrickson, B. L., & Levenson, R. W. (1998). Positive emotions speed recovery from the cardiovascular sequelae of negative emotions. *Cognition and Emotion, 12*(2), 191–220.

Fredrickson, B. L., Mancuso, R. A., Branigan, C., & Tugade, M. M. (2000). The undoing effect of positive emotions. *Motivation and Emotion, 24*(4), 237–258.

Gottman, J. M., & Krokoff, L. J. (1989). The relationship between marital interaction and marital satisfaction: A longitudinal view. *Journal of Consulting and Clinical Psychology, 57,* 47–52.

Gottman, J. M., & Levenson, R. W. (1985). A valid procedure for obtaining self-report of affect in marital interaction. *Journal of Consulting and Clinical Psychology, 53*(2), 151–160.

Gottman, J. M., & Levenson, R. W. (1986). Assessing the role of emotion in marriage. *Behavioral Assessment, 8,* 31–48.

Gottman, J. M., & Levenson, R. W. (1992). Marital processes predictive of later dissolution: Behavior, physiology, and health. *Journal of Personality and Social Psychology, 63*(2), 221–233.

Gottman, J., Notarius, C., Markman, H., Bank, C., Yoppi, B., & Rubin, M. E. (1976). Behavior exchange theory and marital decision making. *Journal of Personality and Social Psychology, 34,* 14–23.

Gottman, J. M., & Ringland, J. T. (1981). The analysis of dominance and bidirectionality in social development. *Child Development, 52,* 393–412.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1995). *International Affective Picture System (IAPS): Technical manual and affective ratings.* Gainesville: University of Florida, Center for Research in Psychophysiology.

Lang, P. J., Levin, D. N., Miller, G. A., & Kozak, M. J.(1983). Fear behavior, fear imagery, and the psychophysiology of emotion: The problem of affective-response integration. *Journal of Abnormal Psychology, 92,* 276–306.

Larson, C. L., Sutton, S. K., & Davidson, R. J. (1998). Affective style, frontal EEG asymmetry and the time course of the emotion-modulated startle. *Psychophysiology, 35,* S52.

Levenson, R. W., Carstensen, L. L., & Gottman, J. M. (1994). The influence of age and gender on affect, physiology, and their interrelations: A study of long-term marriages. *Journal of Personality and Social Psychology, 67*(1), 56–68.

Levenson, R. W., & Gottman, J. M. (1983). Marital interaction: Physiological linkage and affective exchange. *Journal of Personality and Social Psychology, 45*(3), 587–597.

Levenson, R. W., & Gottman, J. M. (1985). Physiological and affective predictors of change in relationship satisfaction. *Journal of Personality and Social Psychology, 49*(1), 85–94.

Levenson, R. W., Oyama, O. N., & Meek, P. S. (1987). Greater reinforcement from alcohol for those at risk: Parental risk, personality risk, and sex. *Journal of Abnormal Psychology, 96*(3), 242–253.

Levenson, R. W., & Ruef, A. M. (1992). Empathy: A physiological substrate. *Journal of Personality and Social Psychology, 63*(2), 234–246.

Levenson, R. W., Sher, K. J., Grossman, L. M., Newman, J., & Newlin, D. B. (1980). Alcohol and stress response dampening: Pharmacological effects, expectancy, and tension reduction. *Journal of Abnormal Psychology, 89,* 528–538.

Markman, H. J. (1979). Application of a behavioral model of marriage in predicting relationship satisfaction of couples planning marriage. *Journal of Consulting and Clinical Psychology, 47,* 743–749.

Markman, H. J. (1981). Prediction of marital distress: A 5-year follow-up. *Journal of Consulting and Clinical Psychology, 49,* 760–762.

Mauss, I. B., Levenson, R. W., McCarter, L. M., Wilhem, F. W., & Gross, J. J. (2005). The tie that binds? Coherence among emotion experience, behavior, and physiology. *Emotion, 5,* 175–190.

Mayer, J. D., & Stevens, A. (1994). An emerging understanding of the reflective (meta-) experience of mood. *Journal of Research in Personality, 28,* 351–373.

Pitman, R. K., & Orr, S. P. (1986). Test of the conditioning model of neurosis: Differential aversive conditioning of angry and neutral facial expressions in anxiety disorder patients. *Journal of Abnormal Psychology, 95,* 208–213.

Pitman, R. K., Orr, S. P., Forgue, D. F., de Jong, J. B., & Claiborn, J. M. (1987). Psychophysiologic assessment of posttraumatic stress disorder imagery in Vietnam combat veterans. *Archives of General Psychiatry, 44,* 970–975.

Ruef, A. M. (2001). Empathy in long-term marriage: Behavioral and physiological correlates (Doctoral dissertation, University of California, Berkeley, 2000). *Dissertation Abstracts International, 62*(01), 563B.

Salovey, P., Mayer, J. D., Goldman, S., Turvey, C., & Palfai, T. (1995). Emotional attention, clarity, and repair: Exploring emotional intelligence using the Trait Meta-Mood Scale. In J. W. Pennebaker (Ed.), *Emotion, disclosure and health* (pp. 125–154). Washington, DC: American Psychological Association.

Sher, K. J., & Levenson, R. W. (1982). Risk for alcoholism and individual differences in the stress-response-dampening effect of alcohol. *Journal of Abnormal Psychology, 91,* 350–367.

Tsai, J. L., & Levenson, R. W. (1997). Cultural influences on emotional responding: Chinese and European American dating couples during interpersonal conflict. *Journal of Cross-Cultural Psychology, 28*(5), 600–625.

Tsai, J. L., Levenson, R. W., & Carstensen, L. L. (2000). Autonomic, subjective, and expressive responses to emotional films in older and younger Chinese Americans and European Americans. *Psychology and Aging, 15*(4), 684–693.

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.