Journal of Personality and Social Psychology
1992, Vol. 63, No. 2, 234–246

0022-3514/92/$3.00

Empathy: A Physiological Substrate

Robert W. Levenson and Anna M. Ruef
University of California, Berkeley

The relation between empathy (defined as the ability to perceive accurately how another person is feeling) and physiology was studied in 31 Ss. Ss viewed 15-min marital interactions and used a rating dial to indicate continuously how they thought a designated spouse was feeling. Rating accuracy was determined by comparing Ss' ratings with identical self-ratings obtained previously from the target spouse. Physiological linkage between S and target was determined using bivariate time-series analyses applied to 5 autonomic and somatic measures obtained from the S during the rating task and from the target spouse during the original conversation. Accuracy of rating negative emotion was greatest when S and target evidenced high levels of physiological linkage across time. Accuracy of detecting positive emotion was related to a state of low cardiovascular arousal in the S, but not to physiological linkage between S and target.

Empathy is a fundamental part of the social fabric of emotion, providing a bridge between the feelings of one person and those of another. The notion that empathy between two people is related to a state of shared physiology is intriguing, suggesting a parallelism between psychological and physiological domains. The study of empathy has generated a large literature, often including physiological measurement, but almost always from a single person (either the person being observed or the observer, but not both). Shared physiology has generated a relatively small literature that has not addressed its relation with empathy.

Empathy: Definition and Measurement

Definitional Issues

The experimental and theoretical literature on empathy has failed to agree on a single definition (Eisenberg & Miller, 1987; Wispé, 1986). In this literature, the term *empathy* appears to have been used to refer to at least three different qualities: (a) *knowing* what another person is feeling (e.g., Dymond, 1949; a component of "content accuracy" in Ickes, Stinson, Bissonnette, & Garcia, 1990[1]); (b) *feeling* what another person is feeling (e.g., Eisenberg, Fabes, Schaller, & Miller, 1989; Feshbach, 1975; Feshbach & Roe, 1968; Mehrabian & Epstein, 1972; Stotland, Matthews, Sherman, Hansson, & Richardson, 1978); and (c) *responding compassionately* to another person's distress (e.g., Batson, O'Quin, Fultz, Vanderplas, & Isen, 1983; Coke, Batson, & McDavis, 1978).

Failure to distinguish between *empathy* and *sympathy* (i.e., compassion or concern stimulated by the distress of another, Gruen & Mendelsohn, 1986) has been noted by several authors (Gladstein, 1984; Hickson, 1985; Wispé, 1986). Reflecting this confusion, some of the characteristics said to typify the empathic person—such as patience, affiliation, liberalism, and humanism (Greif & Hogan, 1973), or warmth, understanding, and openness (Kagan & Schneider, 1987)—clearly represent the most global and nonspecific positive human qualities.

We believe that the most useful definition of empathy would emphasize the ability to detect accurately the emotional information being transmitted by another person. This entity has been termed *empathic accuracy* by others (see Ickes et al., 1990, for a brief review). Related definitions are found in Wispé (1986), who defined empathy as "the attempt by one self-aware self to comprehend unjudgmentally the positive and negative experiences of another self" (p. 318), and in the infancy literature, where empathy has been viewed as the ability to share and understand the child's emotions and signals (e.g., Ainsworth, 1973; Ainsworth, Bell, & Stayton, 1974; Wiesenfeld, Whitman, & Malatesta, 1984). Related constructs include *affective egocentrism* (Ford, 1979) and *empathic perspective taking* (Underwood & Moore, 1982).

Measurement

Given this diversity of definition, it is not surprising that a number of different self-report measures of empathy have emerged (e.g., Dymond, 1949; Hogan, 1969; Mehrabian & Epstein, 1972). These measures have proved to have low intercorrelations (e.g., Kagan & Schneider, 1987; Kurtz & Grummon, 1972), low reliability (e.g., criticisms of the Hogan scale by Cross & Sharpley, 1982), low validity in the form of inconsistent relationships with external criteria (e.g., Deutsch & Madle, 1975), and numerous other problems (for critical reviews see Eisenberg & Lennon, 1983; Wispé, 1986).

This research was supported by National Institute on Alcohol Abuse and Alcoholism Grant AA05004, National Institute of Mental Health Grant MH39895, and National Institute on Aging Grant AG07476 to Robert W. Levenson.

Correspondence concerning this article should be addressed to Robert W. Levenson, Department of Psychology, University of California, Berkeley, California 94720.

[1] In addition to knowledge of what another person is feeling (both its specific content and valence), Ickes, Stinson, Bissonnette, and Garcia (1990) included knowledge of that person's specific thoughts as a component of empathic accuracy.

Conclusions

Considering these conceptual issues of definition and methodological issues of operationalization and measurement, we draw three conclusions. First, the ability to perceive accurately the feelings of another person is arguably the most fundamental aspect of empathy. Without accurate perception of another's feelings, it would be difficult to feel what others feel or to respond compassionately to their plight. Second, to be maximally useful, the construct of empathy should be operationalized in terms of measurable skills, knowledge, and behavior rather than in terms of elusive qualities of "goodness." Third, self-report measures of empathy are particularly vulnerable to distortions due to social desirability (e.g., does anyone want to portray oneself as nonempathic?) and inaccurate self-evaluations (i.e., how do we come to know how empathic we really are?).

Empathy: Physiological Correlates

Empathy and Observer's Physiology

There is evidence that observing a person in distress can produce signs of emotional arousal (i.e., autonomic nervous system, facial expressive, and subjective responses) in the observer (Eisenberg et al., 1988; Eisenberg, Fabes, Miller, et al., 1989; Stotland, 1969; Wiesenfeld, Whitman, & Malatesta, 1984).

There is related evidence that observing the emotional displays of another person can result in similar emotional displays as well as autonomic arousal on the part of the observer (Dimberg, 1982; Lanzetta & Englis, 1989; McHugo, Lanzetta, Sullivan, Masters, & Englis, 1985; Vaughan & Lanzetta, 1980).

Accuracy of Rating Emotion and Target's Physiology

None of the foregoing studies was concerned with the accuracy of emotion rating, which would require instructing subjects to rate the targets' emotions and providing an independent measure of the accuracy of these ratings. In addition, all of these studies focused on the physiological responses of the observer. In contrast, studies using the *sender–receiver* paradigm have been focused on the relation between physiological responses of the target and the extent to which that target's affect could be accurately rated. In general, these studies have found that high physiological arousal on the part of the target is associated with low accuracy on the part of the subject in rating the target's affect (Buck, Savin, Miller, & Caul, 1972; Buck, Miller, & Caul, 1974[2]; Lanzetta & Kleck, 1970).

Buck et al. (1972) also provided one of the few tests of the relation between subject and target physiology in the literatures relevant to empathy. Using simple correlations, subject's and target's physiological responses were found to be uncorrelated overall. No attempt was made to determine whether differences in the extent of this correlation were related to differences in rating accuracy.

Shared Physiology

Whereas the essence of empathy is interpersonal, physiology is almost always viewed as a private, intrapersonal phenomenon. Nonetheless, physiological responses of an individual can often be best understood in terms of basic social processes (e.g., Cacioppo & Petty, 1983), and furthermore, the physiological responses of two people can evidence considerable relatedness and linkage.

Studies of shared physiology have largely been conducted outside of the empathy arena. Linkage between the physiological responses of two people was first demonstrated in studies of psychotherapy and psychodiagnosis (DiMascio, Boyd, & Greenblatt, 1957; DiMascio, Boyd, Greenblatt, & Solomon, 1955; Malmo, Boag, & Smith, 1957). In Kaplan and Bloom's (1960) review of this largely descriptive literature, they interpreted this linkage as indicating a physiological component of empathy. Ax (1964) sounded a similar theme, suggesting that empathy might be thought of "as an autonomic nervous system state which tends to simulate that of another person" (p. 12). Later studies examined shared physiology in relation to more enduring characteristics of the relationship. Kaplan, Burch, and Bloom (1964) found that correlation among subjects' skin conductance increased to the extent that group members disliked each other.

Studies of shared physiology essentially disappeared for 20 years, until we began studying physiological linkage between spouses during marital interaction (Levenson & Gottman, 1983). In this work we tried to rectify several methodological problems that had plagued the early studies. Whereas most earlier studies attempted to characterize shared physiology using a single autonomic nervous system measure, we used a more broadly based set of measures. In addition, the early studies often did not use statistical tests or, if they did, used running correlations to assess shared physiology. Because physiological measures obtained at different times from the same person or from two persons in an interaction are not independent observations, and because autocorrelations (i.e., cyclicity)[3] within the physiological responses of each person can unduly inflate correlations, we quantified shared physiology using a bivariate

---

[2] A closer examination of these studies (Buck, Miller, & Caul, 1974; Buck, Savin, Miller, & Caul, 1972) reveals a more complex picture. Although these studies are typically cited as indicating that high physiological arousal on the part of the sender was associated with low rating accuracy (the finding emphasized by Buck et al. as well), Buck et al. (1972) only found this relation when accuracy was defined in terms of the correlation between observers' and senders' "pleasantness" ratings (and then only for women). The other major finding was entirely different. When accuracy was defined in terms of observers' ability to indicate correctly the category of slide the sender was viewing, high physiological arousal on the part of the observer was associated with high rating accuracy (and then only for men). Unfortunately, in a subsequent study using this paradigm (Buck, Miller, & Caul, 1974), the correlation between this important measure of "categorical" accuracy and subjects' physiological levels was not reported.

[3] As a concrete example of this autocorrelation problem, consider the possibility that two subjects happen to be breathing at approximately the same rate for reasons unrelated to the experimental manipulation. This similar breathing rate should produce in both subjects similar cyclical patterns of rise and fall in heart rate (due to respiratory sinus arrhythmia), which would increase the magnitude of running correlations computed for their heart rates.

time series analysis that was less vulnerable to these problems (Gottman, 1981; Gottman & Ringland, 1981).

Using these methods, we found that physiological linkage during attempted resolution of marital conflicts was greater for unhappily married couples than for happily married couples (Levenson & Gottman, 1983). We interpreted this finding by positing that similarity in patterns of autonomic activation can result from similarities in patterns of negative affect, which should be most prevalent in unhappy marriages and among interactants who dislike each other (e.g., Kaplan et al., 1964). In a subsequent report, we found strong linkage between the physiological responses a spouse had evidenced during a marital interaction and the physiological responses that same spouse evidenced several days later while viewing and rating a videotape of that interaction (Gottman & Levenson, 1985). We interpreted this finding as suggesting that subjects "relived" the experience of the interaction when viewing and rating it at the later date, thus adding credence to the validity of obtaining affect ratings in this manner.

A Behavioral Test for Assessing the Relation Between Empathy and Physiological Linkage

To measure subjects' ability to perceive accurately the feelings of another person, we adapted the videotape viewing and rating procedures used in our studies of marital interaction (e.g., Levenson & Gottman, 1983) to produce a new behavioral measure of empathy.

We needed an experimental context for the behavioral assessment of empathy that would avoid some of the violations of ecological validity found in earlier research. Such a context would have a number of features: (a) Subjects should attempt to judge continuous streams of behavior, rather than summaries or short excerpted episodes; (b) Subjects should have available both visual and auditory information, rather than solely using transcripts, silent videotapes, or audiotapes; and (c) The behavior to be judged should be naturalistic, rather than staged.[4] We also thought it important to separate ability to detect accurately positive and negative affect, reasoning that the ability to detect positive emotion might be somewhat independent from the ability to detect negative emotion and that each type of emotion could bear a different relation with physiology.

Thus, we developed a behavioral assessment procedure in which subjects rated the affects being experienced by a target person who was engaged in a naturalistic interaction with his or her spouse. The subject's rating of the target's affective state would be compared with that target's own affective ratings to determine rating accuracy.[5]

Similarly, the subject's physiology, measured while he or she was rating the target spouse's emotions, would be compared with the target's physiology, measured during the marital interaction. We hypothesized that empathy is associated with a state of shared physiology, such that when one person was most empathically (i.e., accurately) perceiving the feelings of another, the two would most likely be in a common physiological state.

Method

Subjects

Thirty-one subjects (14 men and 17 women) were recruited using advertisements in Berkeley, California area newspapers. Subjects had to be married, be over the age of 21, and have no major physical or mental health problems. Because this study was conducted as part of an ongoing study of alcohol and emotion, subjects were screened to ensure that they were moderate social drinkers with no history of alcoholism. Subjects were paid $10 for completing a package of questionnaires and $8 per hour for the laboratory session. To control for menstrual phase variations, female subjects were scheduled for laboratory sessions between the fifth and ninth day after the start of their most recent menstrual period.

Apparatus and Materials

Questionnaires. Before the laboratory sessions, subjects completed a number of questionnaires, including two measures of marital satisfaction (Burgess, Locke, & Thomes, 1971; Locke & Wallace, 1959) and two self-report measures of empathy. One of these empathy measures was a modification of the Mehrabian and Epstein scale (1972; items were answered true or false instead of on the original 9-point scale) and the other was the Empathy subscale from the California Personality Inventory (CPI; Gough, 1987), which was adapted from Hogan's (1969) measure of empathy.

Stimulus tapes of marital interactions. We initially selected 22 video recordings of 15-min conversations between married spouses from recordings obtained in previous studies of marital interaction conducted in this laboratory. This initial selection was based on the couple having given permission for the tapes of their interactions to be shown to

[4] Although several behavioral tests relevant to empathy exist, none fulfilled all of our criteria. Campbell, Kagan, and Krathwohl's (1971) measure of affective sensitivity probably came closest to meeting our needs in that accuracy was based in part on the relation between subject's and target's emotional ratings. In this method, subjects viewed an excerpt from a counseling session and then selected one of several alternatives that indicated what the client was feeling. The "correct" response was determined by considering both clients' own recall of their affect and the judgment of expert observers. This use of forced-choice ratings and discrete, as opposed to continuous ratings, however, made it unsuitable for our purposes. Other existing behavioral measures of the ability to detect emotion had more serious problems in terms of our requirements. Ekman, Friesen, & O'Sullivan's (1977) Reading Facial Expressions is one of several tests in which subjects attempt to judge the emotions portrayed in pictures of facial expressions (problems: static stimulus, discrete ratings, and not social interaction). Buck's (1976) Communication of Affect Receiving Ability Test (CARAT) has subjects view silent videotaped sequences of people watching slides and guess whether the slide content is sexual, scenic, unpleasant, or unusual (problems: discrete ratings, limited range of emotions, and not social interaction). Rosenthal, Hall, DiMatteo, Rogers, and Archer's (1979) Profile of Nonverbal Sensitivity (PONS) has subjects view 2-s video versions and audio versions of one person posing different emotional responses (problems: discrete ratings, posed stimulus, and not social interaction). Archer and Akert's (1977) Social Interpretation Task (SIT) has subjects view videotaped segments of naturally occurring situations and answer factual questions (problems: discrete ratings and not related to emotion).

[5] Some time after our work was completed, we became aware of Ickes, Stinson, Bissonnette, and Garcia's (1990) method for assessing empathic accuracy, which bears several similarities to ours. In their method, strangers interacted for 6 min, viewed a videotape of the interaction, and wrote down all of their thoughts and feelings (assigning each a positive, negative, or neutral affective valence score), then viewed the videotape again and wrote down all of the thoughts and feelings they thought the other person was having. Empathic accuracy was determined from the agreement between the content and valence information. No physiological measures were obtained.

others in subsequent research, the video and audio recording being of high quality, and our impression that the interactions encompassed a fairly broad range of emotions.

In the studies from which these tapes were culled (see Gottman & Levenson, in press; Levenson & Gottman, 1983, 1985, for full details), married couples came to the laboratory after 8 hr of separation and engaged in three 15-min conversations (each immediately preceded by a 5-min silent preinteraction period). Two of these conversations were used for the present study; in one, the couple either discussed the events of the day or the events of the past 3 years, and, in the other, they attempted to resolve a problem area in their marriage. The conversations were videotaped, and continuous second-by-second averages were obtained for five physiological variables (heart rate, skin conductance, pulse transmission time to the finger, finger pulse amplitude, and somatic activity) from each spouse during the 5-min preinteraction period and during the 15-min conversation.

Several days later, each spouse returned to the laboratory separately and viewed the videotape of the couple's interaction. Each spouse provided a continuous rating of how he or she was feeling during the interaction using a joystick device that traversed a 180° arc over a 9-point scale anchored with the legends *very negative* and *very positive*, with *neutral* at the midpoint. Extensive data have been presented that support the validity of this procedure for assessing affective state (Gottman & Levenson, 1985).

Although it is arguably the case that the only person who truly knows how he or she feels is that person, we still wanted to eliminate interactions in which the target spouse's own affect ratings were so idiosyncratic that there would be no agreement with ratings obtained from independent observers. Thus, the 22 taped conversations were rated by a group of 34 subjects using the same rating dial procedure used in the present study (see description later in this article). These subjects were primarily university staff and students not affiliated with our laboratory. Using these data, we selected four conversations for which there was general agreement between the emotional ratings provided by the target and those provided by the 34 raters. Agreement was determined by examining similarity between targets' and subjects' ratings in terms of mean affect ratings over the 15-min conversation and similarity in second-to-second variation (using a simple index based on the mean square differences). The final set of four conversations was selected from four different couples; in two, the husband was the target spouse, and in the other two, the wife was the target spouse. Two of the conversations were discussions of events of the day and two were discussions of marital problem areas.

Video. Subjects were informed that a video recording would be made of them during the experiment using a partially concealed camera. Subjects viewed the videotapes of the marital interactions on the same 13" color monitor that had been used to show these recordings to the target spouses when their ratings were obtained in the original studies.

Rating dial. The same rating dial device (described above) that had been used to obtain affect ratings from the target spouses in the original study was used to obtain affect ratings from the subjects in the present study. The dial pointer was attached to a potentiometer in a voltage-dividing circuit that provided a signal to the computer system (see below) from which the precise dial position could be determined.

Physiological. While subjects viewed and rated the recorded conversations, the same five[6] physiological functions that had been obtained from the target spouses were measured. A system consisting of two Lafayette Instruments six-channel polygraphs and a DEC LSI 11/73 microcomputer was used to measure: (a) heart rate—miniature electrodes with Redux paste were placed in a bipolar configuration on opposite sides of the subject's chest; (b) skin conductance level—a constant voltage device passed a small voltage between Beckman regular electrodes attached to the palmar surface of the middle phalanges of the first and third fingers of the nondominant hand using an electrolyte of sodium chloride in Unibase; (c) general somatic activity—an electromechanical transducer attached to a platform under the subject's chair generated an electrical signal proportional to the amount of movement in any direction; (d) pulse transmission time to the finger—a UFI photoplethysmograph was attached to the second finger of the nondominant hand. The interval was timed between the R-wave of the EKG and the upstroke of the finger pulse; and (e) finger pulse amplitude—the trough-to-peak amplitude of the finger pulse was measured.

This set of physiological measures was selected to sample broadly from major organ systems (cardiac, vascular, electrodermal, and somatic muscle); to allow for continuous measurement; and to be as unobtrusive as possible. The resolution of the computer/polygraph system was 1 ms for measures of time and 1 mV for measures of amplitude.

Procedure

Subjects arrived at the laboratory and were seated in a chair as we attached the recording devices and explained their functions. The use of the rating dial was explained and subjects were instructed to adjust it as often as necessary so that it always reflected how they thought the target spouse was feeling during the interaction.

The subject was told which spouse to rate during the first conversation and then viewed the video recording of that conversation. The entire 20-min recording was rated (5-min silent preinteraction period followed by 15-min conversation), but only the 15-min conversation was considered in the determination of rating accuracy and physiological linkage. After the tape had been viewed, subjects were asked to indicate (on scales from 1 to 10) how difficult they found the task and how accurate they thought they had been in rating the target spouse's affect. After a 5-min rest period, the same procedure was followed for the second tape. From the set of four stimulus tapes, each subject viewed one recording in which a husband was the target and one in which a wife was the target. Viewing order was counterbalanced between subjects.

After viewing and rating these two tapes, subjects participated in the second part of the experiment, which involved alcohol consumption. Data from this second part of the experiment will not be reported here.

Results

Data Reduction

Physiological means and standard deviations. The computer that processed the physiological data on-line was programmed to compute second-by-second averages for each physiological measure. Using these data, overall means and standard deviations were determined for each measure for the 5-min preinteraction period and for the 15-min interaction period. Raw score averages were also computed for each 10-s period during the 15-min interaction. The z-scores were then computed for each of these 10-s periods on the basis of the means and standard deviations for the 5-min preinteraction silent period. The same data reduction procedures were also carried out for the physio-

[6] Measures of finger temperature, respiration rate, respiration depth, and pulse transmission time to the ear were also obtained; these will not be reported because comparable data were not available from the target spouses.

logical data that had been collected from the target spouses when they had been in the original interactions.[7]

Physiological linkage. We used the 10-s period z score averages from the 15-min interaction in bivariate time-series analyses to determine the extent of *physiological linkage* between the subject and the target using methods based on Gottman (1981) and described in Levenson and Gottman (1983).[8]

This bivariate time-series analysis yielded 10 log-likelihood statistics that have approximately chi-square distributions, two for each of the five physiological variables obtained from the subject and the target. For each physiological variable, one of these statistics represented the extent to which the subject's pattern of response accounted for variance in the target's pattern of response beyond the variance accounted for by the target's autocorrelation; the other represented the extent to which the target's pattern of response accounted for variance in the subject's pattern of response beyond the variance accounted for by the subject's autocorrelation. For each subject, the percentage of the 10 linkage statistics that were statistically significant at the .05 level was determined. The average subject showed significant physiological linkage with the target for 33% (SE = 3%) of the physiological variables in the first tape viewed and 28% (SE = 2%) of the variables during the second tape viewed.

These 10 log-likelihood statistics were converted to z scores. A simple average of 10 z scores was used as a single overall index of the extent of physiological linkage between subject and target. This index of linkage was used in essentially all of the data analyses involving physiological linkage. The one exception was those analyses that attempted to determine whether linkage in *particular* physiological measures was uniquely related to rating accuracy; in that case the 10 individual z scores were used.

Rating accuracy. To determine how accurately subjects rated the target's affect, we adapted a lag sequential analysis that had been used in our earlier work to study "affect reciprocity" across spouses (Levenson & Gottman, 1983).

Each 10-s period of rating dial data during the 15-min interaction was classified as positive, negative, or neutral; this was done separately for subject and target. To be coded positive, the raw score average had to be greater than or equal to 6.0 (referenced to the original 1–9 affect rating dial scales) and the z score had to be greater than or equal to 0.5. Thus, a positive classification meant that, for that period, the pointer was actually on the positive portion of the dial (the raw score criterion) *and* was positive *relative* to the subject's range of ratings (the z-score criterion). To be coded negative, 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.

Rating accuracy scores were determined separately for positive and negative affect at *lag zero* (i.e., target and subject gave same rating in the same 10-s period) and *lag one* (i.e., target's rating in a given 10-s period was matched by the subject's rating in the following period). An accuracy score was also determined for *lag minus one* (i.e., target's rating in a given 10-s period was matched by the subject's rating in the previous 10-s period). Because this lag minus one score reflected accuracy of ratings made by subjects of segments they had not yet seen, it enabled evaluation of the necessity of actually viewing the interaction for making accurate ratings. It should be noted that a match between subject and target ratings for any given 10-s period required both the subject and the target to have rated the period positive or both to have rated it negative; neutral ratings were not counted as matching either positive or negative ratings.

Each rating accuracy z score was determined by subtracting the unconditional probability from the conditional probability and dividing by an estimate of the standard error. For example, the formula for accuracy of rating negative affect at lag one was

$$\frac{(\text{TSNEG1}/\text{TNEG}) - (\text{TNEG}/90)}{\text{SQRT}\{(\text{TNEG}/90) \times (1 - [\text{TNEG}/90])/\text{TSNEG1}\}},$$

where TSNEG1 = number of times in which the target's affect rating was negative in a given period and the subject's affect rating was negative in the following period; TNEG = number of periods in which the target's affect rating was negative; and SQRT = square root.

The interested reader is referred to Levenson and Gottman (1983) for additional detail and to Allison and Liker (1982) for a discussion of various z-score indices of sequential probability.

Affective Characteristics of Stimulus Tapes

The four stimulus tapes used in the rating task were typical of marital discussions in which spouses recount events or attempt to resolve a marital conflict, insofar as the overall affect was skewed toward the negative. In terms of the affect ratings originally given by the target spouse, the average rating of the four tapes for the 15-min interaction was 4.0 (1–9 scale, with 5 representing *neutral*). In terms of the criteria for negative and positive 10-s periods presented in the preceding section, the tapes had an average of 36 negative periods and 7 positive periods (90 periods total).

Overall Empathic Accuracy

Before attempting to account for variation across subjects in the ability to rate accurately targets' affects, we wanted to estimate the mean overall level of rating accuracy. As a simple index, we computed the percentage of 10-s periods rated positive by the target that were also rated positive by the subject in the same rating period. A similar measure was derived for subjects' matching targets' ratings from the previous rating period. This procedure was repeated eight times (positive or negative affect × first or second conversation rated × lag zero or lag one). Examination of Table 1 reveals that the mean level of this accuracy index ranged from 28% to 43% (with six of the eight accuracy percentages larger than the 33% chance level). More important, Table 1 reveals that the performance of individual subjects varied widely (ranging from 0% to 100% accuracy). Thus, it

[7] The p = .05 rejection level was used unless otherwise indicated. All reported significance tests for correlations and t tests were conducted two-tailed.

[8] Conceptually, this analysis first attempts to account for as much of the variance in a given series (e.g., subject'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., target'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).

Table 1
Overall Empathic Accuracy

Empathy variable	Percentage of time target's affect rating is matched by subject's M	SE	Range
Negative affect (first conversation)			
Lag zero	28	4	0–73
Lag one	32	4	0–79
Positive affect (first conversation)			
Lag zero	40	8	0–100
Lag one	43	8	0–100
Negative affect (second conversation)			
Lag zero	34	5	0–79
Lag one	36	5	0–88
Positive affect (second conversation)			
Lag zero	37	7	0–100
Lag one	38	8	0–92

Note. Lag zero = target and subject gave same rating in the same 10-s period; lag one = target's rating in a given 10-s period was matched by the subject's rating in the following period.

appears that this task for assessing empathy had the desirable properties of being neither too simple nor too difficult and of producing a wide range of performances.

Relations Between Physiological Linkage and Empathy

Table 2 presents the correlations between the empathy variables and the overall index of physiological linkage. These revealed that the greater the physiological linkage between subject and target, the greater the accuracy of the subject's rating of the target's negative affect. This finding was obtained for both the first conversation that was rated and for the second conversation, both at lag zero and at lag one.[9] There was no relation between physiological linkage and the accuracy of subjects' rating of positive affect.

To ensure that rating accuracy was in fact based on actually viewing the interaction rather than on some spurious relation between the way that subjects and targets used the rating dial, we looked at the relation between physiological linkage and rating accuracy at lag minus one (subject matches target's rating from the following 10-s period). If this were significant it would indicate that subjects were able to predict accurately the affect in a given 10-s segment before actually seeing that segment. As would be expected, this correlation was not significant in either conversation for either negative or positive affect.

Are Relations Between Physiological Linkage and Empathy More Consistent for Certain Physiological Variables Than for Others?

In the foregoing analyses, a strong relation was found between rating accuracy for negative affect and an index of physiological linkage that encompassed all measured physiological variables. Our previous work studying the relation between physiological linkage and marital satisfaction was limited to the use of this overall index. However, for the present study we wanted to determine whether the relation between rating accuracy and linkage was stronger overall for some physiological variables than for others. Analyzing the 10 linkage z scores for individual variables, we found that, in both the first and second conversations, only linkage in skin conductance and pulse transmission time to the finger were univariately correlated with rating accuracy for negative affect. (First conversation: linkage in skin conductance and rating accuracy at lag zero, $r = .53$, $p = .002$, at lag one, $r = .55$, $p = .002$; linkage in pulse transmission time and rating accuracy at lag one, $r = .36$, $p = .049$. Second conversation: linkage in skin conductance and rating accuracy at lag zero, $r = .63$, $p < .001$, at lag one, $r = .62$, $p < .001$; linkage in pulse transmission time and rating accuracy at lag zero, $r = .54$, $p = .002$, at lag one, $r = .49$, $p = .005$.) Correlations between rating accuracy for negative affect and linkage in individual physiological variables are presented in Table 3.

Can Rating Accuracy Be Predicted From Subjects' Mean Physiological Levels and Variabilities?

Having found that shared physiology between subject and target was associated with greater accuracy in rating negative affect, we found it important to determine whether this same relation would obtain when only the subject's physiology was considered. An analysis of the relation between rating accuracy for negative affect and mean levels and variabilities for each of the five physiological measures averaged over the entire 15-min interaction revealed no relation between any of these mean

[9] To afford some protection against the effects of outliers, correlations between physiological linkage and accuracy of rating negative affect in both conversations were recalculated after removing the most extreme case for each variable. The direction of all correlations was unchanged and all remained statistically significant. To provide an estimate of the "base rate" level of correlation between accuracy of rating negative affect and physiological linkage, a reviewer suggested that we compute the correlation between physiological linkage (for subjects and targets in this sample) and rating accuracy for another group of subjects who viewed the same target tapes. Another interpretation of this correlation is that it provides some control for the possibility that the correlation between rating accuracy and physiological linkage is solely due to some quality of the stimulus tapes (i.e., certain tapes are easier to rate and have the property of provoking the same physiological response in subjects as had been manifested by targets). For this purpose, we used the independent sample of 34 subjects whose ratings had been used to select the set of videotapes that were used in the present study. Fifteen of these subjects had viewed one of the four videotapes used in the present study. The correlations between the accuracy of the ratings for negative affect for these 15 subjects and the physiological linkage between subjects and targets in the present study were nonsignificant for lag zero, $r = -.01$, and for lag one, $r = -.03$. If the logic of this analysis is accepted, then (a) the magnitude of the relationship between accuracy of rating negative affect and physiological linkage found in the present study far exceeded base rate levels; and (b) covariation in the stimulus tapes in terms of ease of rating and capacity to provoke the same autonomic response in subjects as had been manifested by targets was not responsible for our findings.

Table 2
Correlations (N = 31) Between Empathy Variables and Overall Index of Physiological Linkage

Empathic accuracy variable	Correlation with physiological linkage
First conversation	
Negative affect	
Lag zero	.46*
Lag one	.46*
Lag minus one	.27
Positive affect	
Lag zero	.16
Lag one	.16
Lag minus one	−.17
Second conversation	
Negative affect	
Lag zero	.54*
Lag one	.50*
Lag minus one	−.06
Positive affect	
Lag zero	.23
Lag one	.19
Lag minus one	−.11

Note. Lag zero = target and subject gave same rating in the same 10-s period; lag one = target's rating in a given 10-s period was matched by the subject's rating in the following period; lag minus one = target's rating in a given 10-s period was matched by the subject's rating in the previous 10-s period.
* $p < .01$.

levels or variabilities and accuracy in either the first or second conversation. Thus, rating accuracy for negative affect was uniquely associated with measures of shared physiology.

For rating accuracy for positive affect, however, which had been found to be unrelated to physiological linkage, relations were found with mean levels and variability in two cardiovascular measures. In the first conversation, larger pulse amplitudes were related to rating accuracy for positive affect at lag zero, $r = .55$, $p = .002$, and at lag one, $r = .51$, $p = .004$; and greater pulse amplitude variability was related to rating accuracy at lag zero, $r = .57$, $p = .001$, and at lag one, $r = .60$, $p < .001$. In the second conversation, longer interbeat intervals (i.e., slower heart rates) were related to rating accuracy for positive affect at lag one, $r = .36$, $p = .047$. Thus, rating accuracy for positive affect was uniquely associated with measures of subjects' physiology, with the pattern of findings suggesting an association with lower cardiovascular arousal (i.e., greater vasodilation and slower heart rate).

An analysis of rating accuracy in terms of the target's physiological levels was considered inappropriate given that there were only four targets.

Is Rating Accuracy Related to Traditional Empathy Scales?

Rating accuracy was generally unrelated to the traditional empathy scales used. There was no relation between the CPI Empathy subscale and rating accuracy for either positive or negative affect during either the first or second conversation. There was no relation between the Mehrabian–Epstein empathy scale and rating accuracy for either positive or negative affect on the first conversation or for negative affect during the second conversation. The only relation between these traditional scales and rating accuracy was for the second conversation: High scores on the Mehrabian–Epstein empathy scale were associated with greater rating accuracy for positive affect at lag zero, $r = .38$, $p = .033$. Table 4 presents these correlations.

Table 3
Correlations (N = 31) Between Empathy Variables for Negative Affect and Index of Physiological Linkage for Individual Variables

Physiological variable	Lag zero S > T	Lag zero T > S	Lag one S > T	Lag one T > S
First conversation				
Heart rate	−.25	−.05	−.27	−.01
General somatic activity	−.07	−.06	−.05	−.09
Skin conductance	.24	.53**	.21	.55**
Pulse transmission time to the finger	.28	.15	.36*	.29
Finger pulse amplitude	−.22	−.25	−.14	−.20
Second conversation				
Heart rate	−.15	−.19	−.14	−.21
General somatic activity	−.17	−.17	−.10	−.17
Skin conductance	.09	.63**	.06	.62**
Pulse transmission time to the finger	.54**	.05	.49**	.05
Finger pulse amplitude	.07	−.01	.05	−.04

Note. Lag zero = target and subject gave same rating in the same 10-s period; lag one = target's rating in a given 10-s period was matched by the subject's rating in the following period; S > T = subject's physiology predicts target's physiology; T > S = target's physiology predicts subject's physiology.
* $p < .05$. ** $p < .01$.

Table 4
Correlations (N = 31) Between Empathy Variables and Self-Report Measures

	Self-report measure				
Empathy variable	CPI	Mehrabian–Epstein	Difficulty	Accuracy	Marital satisfaction
			Accuracy rating negative affect		
First conversation					
Lag zero	−.35	−.10	−.19	.17	.23
Lag one	−.32	−.26	−.27	.16	.26
Second conversation					
Lag zero	.00	.29	−.42*	.29	−.03
Lag one	−.04	.29	−.41*	.25	−.03
			Accuracy rating positive affect		
First conversation					
Lag zero	−.17	.19	−.07	.26	−.16
Lag one	−.24	.06	−.13	.30	−.13
Second conversation					
Lag zero	−.11	.38*	−.14	.07	−.17
Lag one	−.07	.33	−.16	.17	−.06

Note. Lag zero = target and subject gave same rating in same 10-s period; lag one = target's rating in a given 10-s period was matched by the subject's rating in the following period. CPI = California Personality Inventory.
* $p < .05$.

Is Rating Accuracy Related to Perceived Accuracy, Subject's Marital Satisfaction, or Perceived Difficulty?

Subjects' perceived accuracy of their ratings and their own level of marital satisfaction were unrelated to rating accuracy for either positive or negative affect during either the first or second conversation. There was also no relation between subjects' perceived difficulty of the rating task and rating accuracy for positive or negative affect in the first conversation or for positive affect in the second conversation. Only in the second conversation was any relation found between perceived difficulty and rating accuracy; perceived difficulty was related to low accuracy in detecting negative affect at lag zero, $r = -.42$, $p = .019$, and lag one, $r = -.41$, $p = .024$. Table 4 presents these correlations.

Correlations Between Empathy Scales and Physiological Linkage

The Mehrabian–Epstein and CPI Empathy scales were not significantly correlated with each other ($r = .32$). Neither scale was correlated with physiological linkage during the first conversation (Mehrabian–Epstein $r = -.18$; CPI $r = -.25$) or during the second conversation (Mehrabian–Epstein $r = .03$; CPI $r = .02$).

Sex Differences

There were no significant differences between male and female subjects in (a) the amount of physiological linkage with targets in either conversation, (b) the accuracy of rating negative affect in either conversation, (c) the accuracy of rating positive affect in the first conversation, (d) perceived difficulty of the rating task in either conversation, (e) perceived accuracy on the task in either conversation, and (f) scores on the Mehrabian–Epstein and the CPI Empathy scales.

In the second conversation, women had higher rating accuracy than men for positive affect at lag zero (women's $M = 10.63$; men's $M = 1.58$), $t(18) = -3.07$, $p = .006$, and lag one (women's $M = 11.29$; men's $M = 2.50$), $t(24) = -2.66$, $p = .014$.

Means and tests of significance based on subject sex are presented in Table 5. An analysis of rating accuracy differences in terms of sex of the *target* was considered inappropriate given that there were only two targets of each sex, and their method of selection was such that they could not be considered representative of the general population of male and female targets.

Discussion

The primary hypothesis of this research, that there is a physiological substrate for empathy, was supported. For negative affect, subjects who were best able to rate accurately the negative affect experienced by another person evidenced patterns of physiological response that were similar to those manifested by that person. For positive affect the relation between empathy and physiology involved *individual* rather than *shared* physiology: Subjects best able to rate accurately the positive affects experienced by another person evidenced lower cardiovascular arousal.

Empathy and Shared Physiology

Findings of a relation between greater physiological linkage and greater ability to rate negative affect accurately were quite

Table 5
Sex Differences in Physiological, Affective, and Self-Report Measures

Variable	Men	Women	t
Physiological linkage			
First conversation	8.19	7.41	0.18
Second conversation	4.49	7.37	−0.88
Accuracy rating negative affect			
First conversation			
Lag zero	0.54	0.68	−0.07
Lag one	3.15	1.61	0.53
Second conversation			
Lag zero	1.39	6.12	−1.57
Lag one	1.88	9.13	−1.90
Accuracy rating positive affect			
First conversation			
Lag zero	3.74	7.21	−1.18
Lag one	4.70	7.37	−0.80
Second conversation			
Lag zero	1.58	10.63	−3.07**
Lag one	2.50	11.29	−2.66*
Perceived difficulty			
First conversation	4.18	4.94	−1.06
Second conversation	4.68	5.00	−0.35
Perceived accuracy			
First conversation	6.25	7.06	−1.17
Second conversation	6.86	7.18	−0.49
Mehrabian–Epstein scale	10.14	15.00	−1.64
CPI Empathy scale	24.00	23.06	0.72

Note. Lag zero = target and subject gave same rating in the same 10-s period; lag one = target's rating in a given 10-s period was matched by the subject's rating in the following period; CPI = California Personality Inventory.
* $p < .05$. ** $p < .01$.

consistent. Because the present study used ratings of two conversations, it allowed for an internal replication of findings. The same relation between linkage and rating accuracy for negative affect was found in both conversations, and for both lag zero and lag one (i.e., subjects' ratings showed agreement with targets' ratings from the same 10-s period and from the previous period). And, as should be the case, this relation was *not* found for either conversation at lag minus one, indicating that the relation did not obtain for targets' ratings in periods not yet seen by the subject. Given the robustness of this relation, we offer some speculation as to why it might exist.

Mediation by commonality of emotion. We believe that the relation between physiological linkage and rating accuracy for negative affect is built on commonality of emotional experience. Empathic subjects (i.e., those who are most accurate in rating the negative emotions of targets) would be most likely to experience the same negative emotions, albeit in miniaturized form, at approximately the same time as had the targets. These negative emotions would produce similar patterns of autonomic activation in both subject and target, thus resulting in high levels of physiological linkage. The explicit instruction to rate the feelings of the target person might have increased this contagion of emotion beyond what would have occurred had the subject merely passively viewed a tape of strangers interacting. If this speculation is correct, then our behavioral measure of empathic accuracy taps two of the major characteristics of empathy reviewed in the introduction to this article (i.e., knowing what another person is feeling and feeling what another person is feeling).

Of course, we do not know whether empathic subjects *actually* experienced the same emotions as targets. We do know that they evidenced physiological similarities, but we neither asked subjects how they felt nor compared similarities between the behavioral signs of emotion exhibited by subjects and targets, both of which would have been informative in this regard.

Emotion-specific autonomic activity. For commonality of emotion to result in common patterns of physiological activity, there would need to be some reliable mapping of different physiological patterns onto different emotional states. We have recently presented fairly extensive evidence for the existence of such specificity of autonomic patterning for negative emotions of anger, disgust, fear, and sadness, with indication that this patterning is robust to variations in (a) subject population and experimental procedure (i.e., in actors and college students, regardless of whether subjects have visual emotional information available; Ekman, Levenson, & Friesen, 1983; Levenson, Ekman, & Friesen, 1990); (b) age and mode of elicitation (i.e., in young and in very old subjects, elicited by voluntary facial action or by reliving past emotional experiences; Levenson, Carstensen, Friesen, & Ekman, 1991); and (c) culture (i.e., in Americans and in the Minangkabau of West Sumatra; Levenson, 1989; Levenson, Ekman, Heider, & Friesen, 1992). Levenson (1992) reviewed evidence from other laboratories that supports these autonomic differences among negative emotions.

Although this experiment cannot prove that autonomic patterning is the basis for the found relation between physiological linkage and rating accuracy for negative affect, it seems to us to be a viable and parsimonious explanation for these findings. It is also consistent with interpretations offered in our previous studies of physiological linkage, which were conducted in the context of marital interaction. When we found linkage to be associated with marital distress (Levenson & Gottman, 1983), we stated that we believed "physiological linkage reflects the ebb and flow of negative affect" (p. 596). When we found linkage between subjects' physiological responses when they were interacting with their spouses and when they were viewing and rating video recordings of those interactions (Gottman & Levenson, 1985), we interpreted it in terms of "a distinctive pattern of emotional arousal [that] would produce the same pattern of physiological arousal . . . regardless of whether the emotions were aroused in an interaction or while viewing a videotape of the interaction" (p. 152).

Linkage related to rating accuracy for negative emotion, but not positive emotion. The evidence that emotions produce differentiated autonomic patterning is certainly stronger for negative emotions such as anger and fear than for positive emotions such as joy and relaxation (e.g., Levenson, 1988). In this regard, it is noteworthy that the relation between physiological linkage and empathy found in the present study was very strong for accuracy of rating negative affect, but was totally absent for positive affect. Given our hypothesized explanation for our findings based on commonality of emotional experience, we would have to conclude either that (a) subjects did not experience the positive emotions experienced by targets, (b) subjects experienced the positive emotions at intensity levels too low to recruit autonomic nervous system activity, or (c) subjects experienced the positive emotions, but these positive emotions do not produce the kind of differentiated autonomic activity produced by negative emotions. Given the ready contagion of positive emotion (especially smiles and laughter) and the lack of empirical evidence for patterned autonomic nervous system activity in positive emotion, we would endorse the latter explanation: Most positive emotions do not produce the kinds of patterned autonomic activity necessary for physiological linkage.[10]

The role of sympathetic nervous system activation. Although our major findings were based on a composite measure of physiological linkage that was averaged across all five individual physiological measures, exploratory analyses of the individual measures suggested that it was linkage in skin conductance and pulse transmission time to the finger that were *most consistently* related to rating accuracy for negative emotions. Unlike heart rate, which reflects both sympathetic and parasympathetic influences, these two measures are essentially under exclusive control of the sympathetic branch of the autonomic nervous system. While recognizing the clear limits of the present study in terms of isolating autonomic mechanisms, we should note that these findings are consistent with biological accounts of emotion from Cannon (1927) on, which have given particular emphasis to sympathetic nervous system activation in the service of negative emotions.

An alternative cognitive explanation. We have hypothesized that it is similarity in the ebb and flow of emotion between targets and the subjects who most accurately rate their negative affect that is responsible for similarity in their physiological responses over time. It is of course possible that this physiological linkage resulted from similarity in patterns of *cognitive* activity, especially in those cognitive activities thought to be associated with specific patterns of autonomic nervous system activation (e.g., orienting, attention, and environmental rejection).

Although initially appealing, we consider shared cognitions to be a less viable explanation of our findings than shared emotions. First, the task of the subject (rate the affect of a stranger in a novel situation) and the task of the target (rate one's own affect in a situation in which one was previously a participant) seem quite different cognitively. Second, the affect rating task does not appear to vary greatly over its time course in terms of cognitive demands (e.g., one has to make the same kinds of emotional judgments throughout); without variation in cognitive activity the opportunity for shared variation in associated physiology is lessened. These comments notwithstanding, our experimental design clearly does not enable us to make a definitive choice between emotional and cognitive explanations.

Empathy, Physiology, and Positive Emotion

Our finding that low levels of cardiovascular arousal (i.e., increased peripheral vasodilation and lower heart rate) were associated with greater rating accuracy for positive emotions deserves some comment. However, because these findings were not consistent across both conversations, we will interpret them with some caution.

One possible explanation for this finding is that accurate detection of positive affect in marital interaction may require more extensive and subtle kinds of cognitive activities and judgments than are involved in detection of negative affect. Negative affect in marital interaction is likely to be more obvious, clear-cut, and attention grabbing than positive affect. A great deal of psychophysiological research over the past several decades has investigated the notion that high levels of cardiovascular arousal dampen the activity of higher brain centers by action of the baroreceptor reflex (e.g., Lacey, Kagan, Lacey, & Moss, 1963). Thus, low cardiovascular arousal may contribute to a cognitive state that is more conducive to the greater cognitive demands associated with the detection of positive as compared with negative affect.

Sex Differences

The detection of positive emotion provided the only sex differences we found, with women showing higher accuracy than

[10] We must also consider the possibility that the lack of relation between rating accuracy for positive emotion and physiological linkage resulted from the fact that the stimulus tapes had many more negative affective periods than positive, thus producing a potentially restricted range for positive rating accuracy that could reduce its correlation with other variables. Although we do not completely reject this possibility, we should note that a number of significant relations were found between rating accuracy for positive affect and other variables (i.e., relations with gender, self-report empathy scales, and mean physiological levels)—variables that were unrelated to rating accuracy for negative affect.

men in rating positive affect during the second conversation, a result consistent with findings of women's superiority in the detection of nonverbal cues (e.g., Buck, 1984; Buck, Savin, Miller, & Caul, 1972). Beyond this one difference, we found both sexes to be equally skillful in rating emotion. Although our sample size ($N = 31$) was not large enough for optimal sensitivity, our failure to find pervasive sex differences is consistent with Eisenberg and Lennon's (1983) review of the literature on sex differences in empathy and related abilities. These authors concluded that sex differences are rarely found when measures other than self-report are used. It is also consistent with Ickes et al.'s (1990) finding of no sex differences using a behavioral measure of empathic content accuracy.

Existing Literature on Empathy and Physiology

Our findings weave together and expand on two strands of work reviewed earlier, one finding that viewing the distress and emotional displays of others is physiologically arousing (e.g., Eisenberg et al., 1988, Eisenberg, Fabes, Miller, et al., 1989; Lanzetta & Englis, 1989) and the other finding a relation between physiological arousal on the part of the target and the accuracy with which an observer could detect aspects of that target's emotional state (e.g., Buck et al., 1972, 1974). Using a quite different paradigm, the present study confirms the physiologically arousing qualities of viewing the emotional behavior of others and suggests that aspects of the relation between the subject's and target's physiology, as well as aspects of the subject's physiology, are related to the accuracy of rating both positive and negative emotion.

Self-Report Measures of Empathy

The two self-report measures of empathy used in the present study showed little relation to subjects' ability to rate accurately another person's affect. In this regard, the CPI Empathy subscale was found to be totally unrelated to the accuracy of affect ratings. The variant of the Mehrabian–Epstein scale we used made only mildly superior predictions, predicting rating accuracy for positive affect only, and only in the second conversation. Admittedly, neither of these scales was designed explicitly to measure the ability to discern how another person is feeling. However, to the extent that this ability is accepted as fundamental to empathy, then this lack of relation is a serious problem. Moreover, other self-report measures of empathy, which *are* purportedly related to this ability (Ickes's [1988] measure of empathic accuracy and Davis's [1983] Interpersonal Reactivity Index), also have not fared well in predicting performance on a behavioral measure of empathic accuracy (Ickes et al., 1990). Finally, subjects' self-reports of their own empathic accuracy (Ickes et al., 1990) and of how well they had just done in a test of empathic accuracy (the present study) have been found to be unrelated to their performance on a behavioral test of empathic accuracy.

Empathic Accuracy

A new behavioral measure. For this research, we operationalized empathy in terms of a measurable skill, that of being able to perceive accurately how another person is feeling. By developing a behavioral measure to assess this skill, we avoided a number of the problems inherent in paper-and-pencil measures that either ask subjects how empathic they think they are, or how well they think they can do certain things.

Although the behavioral measure used in the present study certainly is not the only viable way to measure empathy, we believe it assesses a fundamental aspect of empathy (the ability to know how another person is feeling) that is critical to other aspects of empathy such as feeling what another is feeling and responding compassionately to another's distress. Compared with alternative measures, this measure has the advantage of closely approximating the natural context in which empathy occurs, insofar as the rater observes behavior that is naturalistic, complex, continuous, meaningful, and interpersonal and must continuously decode and rate the emotions of another person in real time as the interaction unfolds. Of all the alternative behavioral measures mentioned in the introduction to this article, only the method of Ickes et al. (1990) comes close to meeting these criteria. Our measure, which was useful in assessing ability to detect emotional valence and intensity, is complemented by Ickes et al.'s measure, which was useful for assessing the ability to discern content, but not emotional valence.

Alternative measures. Accuracy on our behavioral measure of empathy was determined by evaluating aspects of the match between the subject's and target's absolute (i.e., raw score criterion) and relative (i.e., z score criterion) valence and intensity ratings, as well as the timing of the ratings (i.e., within the same or adjacent 10-s period).

Our belief in the importance of assessing empathic accuracy in real time limits the amount of information that can be obtained during a single viewing of a tape. Ratings of additional emotional dimensions or of discrete emotions such as anger, fear, and joy could be obtained if subjects were allowed to stop the tape periodically or to view the tape repeatedly. We consider both of these procedures to have problems with external validity (one can rarely stop an interaction or view it repeatedly when making emotional judgments outside the laboratory) and with assessing the relation between empathy and physiology (the interruption method makes temporal comparison of physiology between subject and target extremely difficult; the multiple-pass procedure runs the risk of habituation diminishing emotional and physiological arousal occurring on later passes).

Our analysis of rating accuracy assessed the detection of relatively clear-cut "moments" of positive and negative affect, and not subjects' ability to match every nuance of the target's ratings. In the pilot stages of this work, we evaluated other measures of rating accuracy that took into account more of the available rating dial information. However, few, if any, subjects could accurately track targets' ratings at that level of precision. Thus, it may be that the kind of empathy we have studied here functions primarily to detect the broader strokes of emotion and not the finer details.

Conclusion

These findings are supportive of a physiological substrate for empathy, here defined as the ability to detect accurately how another person is feeling. Accurate rating by one person of another person's positive emotions was associated with a state of low cardiovascular arousal. Accurate rating by one person of another person's negative emotions was associated with a state of shared physiology in which rater and target evidenced similar patterns of autonomic response over time. Although the notion of a shared physiological substrate for empathy was first advanced over 25 years ago (Ax, 1964; Kaplan & Bloom, 1960), we believe this is the first compelling empirical demonstration of this phenomenon.

References

Ainsworth, M. D. S. (1973). The development of infant–mother attachment. In B. M. Caldwell & H. N. Ricciuti (Eds.), *Review of child development research III.* (pp. 1–94). Chicago: University of Chicago Press.

Ainsworth, M. D. S., Bell, S. M., & Stayton, D. J. (1974). Infant–mother attachment and social development: Socialization as a product of reciprocal responsiveness to signals. In M. P. Richards (Ed.), *The integration of a child into a social world.* (pp. 99–136). Cambridge, England: Cambridge University Press.

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

Archer, D., & Akert, R. M. (1977). Words and everything else: Verbal and nonverbal cues in social interpretation. *Journal of Personality and Social Psychology, 35,* 443–449.

Ax, A. A. (1964). Goals and methods of psychophysiology. *Psychophysiology, 1,* 8–25.

Batson, C. D., O'Quin, K., Fultz, J., Vanderplas, M., & Isen, A. M. (1983). Influence of self-reported distress and empathy on egoistic versus altruistic motivation to help. *Journal of Personality and Social Psychology, 45,* 706–718.

Buck, R. (1976). A test of nonverbal receiving ability: Preliminary studies. *Human Communication Research, 2,* 162–171.

Buck, R. (1984). *The communication of emotion.* New York: Guilford Press.

Buck, R., Miller, R. E., & Caul, W. F. (1974). Sex, personality and physiological variables in the communication of emotion via facial expression. *Journal of Personality and Social Psychology, 30,* 587–596.

Buck, R. W., Savin, V. J., Miller, R. E., & Caul, W. F. (1972). Communication of affect through facial expressions in humans. *Journal of Personality and Social Psychology, 23,* 362–371.

Burgess, E. W., Locke, H. J., & Thomes, M. M. (1971). *The family.* New York: Van Nostrand Reinhold.

Cacioppo, J. T., & Petty, R. E. (1983). *Social psychophysiology: A sourcebook.* New York: Guilford Press.

Campbell, R. J., Kagan, N., & Krathwohl, D. R. (1971). The development and validation of a scale to measure affective sensitivity (empathy). *Journal of Counseling Psychology, 18,* 407–412.

Cannon, W. B. (1927). The James–Lange theory of emotions: A critical examination and an alternative theory. *American Journal of Psychology, 39,* 106–124.

Coke, J. S., Batson, C. D., & McDavis, K. (1978). Empathic mediation of helping: A two-stage model. *Journal of Personality and Social Psychology, 36,* 752–766.

Cross, D. G., & Sharpley, C. F. (1982). Measurement of empathy with the Hogan empathy scale. *Psychological Reports, 50,* 62.

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology, 44,* 113–126.

Deutsch, F., & Madle, R. A. (1975). Empathy: Historic and current conceptualizations, measurement, and a cognitive theoretical perspective. *Human Development, 18,* 267–287.

DiMascio, A., Boyd, R. W., & Greenblatt, M. (1957). Physiological correlates of tension and antagonism during psychotherapy: A study of "interpersonal physiology." *Psychosomatic Medicine, 19,* 99–104.

DiMascio, A., Boyd, R. W., Greenblatt, M., & Solomon, H. C. (1955). The psychiatric interview: A sociophysiologic study. *Diseases of the Nervous System, 16,* 4–9.

Dimberg, U. (1982). Facial reactions to facial expressions. *Psychophysiology, 19,* 643–647.

Dymond, R. F. (1949). A scale for the measurement of empathic ability. *Journal of Consulting Psychology, 13,* 127–133.

Eisenberg, N., Fabes, R. A., Miller, P. A., Fultz, J., Shell, R., Mathy, R. M., & Reno, R. R. (1989). Relation of sympathy and personal distress to prosocial behavior: A multimethod study. *Journal of Personality and Social Psychology, 57,* 55–66.

Eisenberg, N., Fabes, R. A., Schaller, M., & Miller, P. A. (1989). Sympathy and personal distress: Development, gender differences, and interrelations of indexes. In N. Eisenberg (Ed.), *Empathy and related emotional responses.* San Francisco: Jossey-Bass.

Eisenberg, N., & Lennon, R. (1983). Sex differences in empathy and related capacities. *Psychological Bulletin, 94,* 100–131.

Eisenberg, N., & Miller, P. A. (1987). The relation of empathy to prosocial and related behaviors. *Psychological Bulletin, 101,* 91–119.

Eisenberg, N., Schaller, M., Fabes, R. A., Bustamante, D., Mathy, R. M., Shell, R., & Rhodes, K. (1988). The differentiation of personal distress and sympathy in children and adults. *Developmental Psychology, 24,* 766–775.

Ekman, P., Friesen, W. V., & O'Sullivan, M. (1977). *Reading facial expressions.* Kenilworth, NJ: Schering Corporation.

Ekman, P., Levenson, R. W., & Friesen, W. V. (1983). Autonomic nervous system activity distinguishes among emotions. *Science, 221,* 1208–1210.

Feshbach, N. D. (1975). Empathy in children: Some theoretical and empirical considerations. *The Counseling Psychologist, 5,* 25–30.

Feshbach, N. D., & Roe, K. (1968). Empathy in six- and seven-year olds. *Child Development, 39,* 133–145.

Ford, M. E. (1979). The construct validity of egocentrism. *Psychological Bulletin, 86,* 1169–1188.

Gladstein, G. A. (1984). The historical roots of contemporary empathy research. *Journal of the History of the Behavioral Sciences, 20,* 38–59.

Gottman, J. M. (1981). *Time-series analysis: A comprehensive introduction for social scientists.* Cambridge, England: Cambridge University Press.

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

Gottman, J. M., & Levenson, R. W. (in press). Toward a typology of marriage based on affective behavior: Preliminary differences in behavior, physiology, health, and risk for dissolution. *Journal of Personality and Social Psychology.*

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

Gough, H. G. (1987). *California psychological inventory administrator's guide.* Palo Alto, CA: Consulting Psychologists Press.

Greif, E. B., & Hogan, R. (1973). The theory and measurement of empathy. *Journal of Counseling Psychology, 20,* 280–284.

Gruen, R. J., & Mendelsohn, G. (1986). Emotional responses to affective displays in others: The distinction between empathy and sympathy. *Journal of Personality and Social Psychology, 51,* 609–614.

Hickson, J. (1985). Psychological research on empathy: In search of an elusive phenomenon. *Psychological Reports, 57,* 91–94.

Hogan, R. (1969). Development of an empathy scale. *Journal of Consulting and Clinical Psychology, 33,* 307–316.

Ickes, W. (1988). *An 18-item measure of empathic accuracy.* Unpublished scale and associated psychometric data, University of Texas at Arlington.

Ickes, W., Stinson, L., Bissonnette, V., & Garcia, S. (1990). Naturalistic social cognition: Empathic accuracy in mixed-sex dyads. *Journal of Personality and Social Psychology, 59,* 730–742.

Kagan, N., & Schneider, J. (1987). Toward the measurement of affective sensitivity. *Journal of Counseling and Development, 65,* 459–464.

Kaplan, H. B., & Bloom, S. W. (1960). The use of sociological and social-psychological concepts in physiological research: A review of selected experimental studies. *Journal of Nervous and Mental Disease, 131,* 128–134.

Kaplan, H. B., Burch, N. R., & Bloom, S. W. (1964). Physiological covariation and sociometric relationships in small peer groups. In P. H. Leiderman & D. Shapiro (Eds.), *Psychobiological approaches to social behavior.* Stanford, CA: Stanford University Press.

Kurtz, R. R., & Grummon, D. L. (1972). Different approaches to the measurement of therapist empathy and their relationship to therapist outcomes. *Journal of Consulting and Clinical Psychology, 39,* 106–115.

Lacey, J. I., Kagan, J., Lacey, B. C., & Moss, H. A. (1963). The visceral level: Situational determinants and behavioral correlates of autonomic response patterns. In P. H. Knapp (Ed.), *Expression of the emotions in man* (pp. 161–196). Madison, CT: International Universities Press.

Lanzetta, J. T., & Englis, B. G. (1989). Expectations of cooperation and competition and their effects on observers' vicarious emotional responses. *Journal of Personality and Social Psychology, 56,* 543–554.

Lanzetta, J. T., & Kleck, R. E. (1970). Encoding and decoding of nonverbal affect in humans. *Journal of Personality and Social Psychology, 16,* 12–19.

Levenson, R. W. (1988). Emotion and the autonomic nervous system: A prospectus for research on autonomic specificity. In H. Wagner (Ed.), *Social psychophysiology and emotion: Theory and clinical applications* (pp. 17–42). New York: Wiley.

Levenson, R. W. (1989, January). *Emotion-specific patterns of autonomic activity are consistent across cultures.* Paper presented at the meeting of the American Association for the Advancement of Science, San Francisco.

Levenson, R. W. (1992). Autonomic nervous system patterning in emotion. *Psychological Science, 3,* 23–27.

Levenson, R. W., Carstensen, L. L., Friesen, W. V., & Ekman, P. (1991). Emotion, physiology, and expression in old age. *Psychology and Aging, 6,* 28–35.

Levenson, R. W., Ekman, P., & Friesen, W. V. (1990). Voluntary facial action generates emotion-specific autonomic nervous system activity. *Psychophysiology, 27,* 363–384.

Levenson, R. W., Ekman, P., Heider, K., & Friesen, W. V. (1992). Emotion and autonomic nervous system activity in the Minangkabau of West Sumatra. *Journal of Personality and Social Psychology, 62,* 972–988.

Levenson, R. W., & Gottman, J. M. (1983). Marital interaction: Physiological linkage and affective exchange. *Journal of Personality and Social Psychology, 45,* 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,* 85–94.

Locke, H. J., & Wallace, K. M. (1959). Short marital-adjustment and prediction tests: Their reliability and validity. *Marriage and Family Living, 21,* 251–255.

Malmo, R. B., Boag, T. J., & Smith, A. A. (1957). Physiological study of personal interaction. *Psychosomatic Medicine, 19,* 105–119.

McHugo, G. J., Lanzetta, J. T., Sullivan, D. G., Masters, R. D., & Englis, B. G. (1985). Emotional reactions to a political leader's expressive displays. *Journal of Personality and Social Psychology, 49,* 1513–1529.

Mehrabian, A., & Epstein, N. (1972). A measure of emotional empathy. *Journal of Personality, 40,* 525–543.

Rosenthal, R., Hall, J. A., DiMatteo, M. R., Rogers, P. L., & Archer, D. (1979). *Sensitivity to nonverbal cues: The PONS test.* Baltimore: Johns Hopkins University Press.

Stotland, E. (1969). Exploratory investigations of empathy. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (pp. 271–314). San Diego, CA: Academic Press.

Stotland, E., Matthews, K. E., Jr., Sherman, S. E., Hansson, R. O., & Richardson, B. Z. (1978). *Empathy, fantasy, and helping.* Beverly Hills, CA: Sage.

Underwood, B., & Moore, B. (1982). Perspective taking and altruism. *Journal of Personality and Social Psychology, 91,* 143–173.

Vaughan, K. B., & Lanzetta, J. T. (1980). Vicarious instigation and conditioning of facial expressive and autonomic responses to a model's expressive display of pain. *Journal of Personality and Social Psychology, 38,* 909–923.

Wiesenfeld, A. R., Whitman, P. B., & Malatesta, C. Z. (1984). Individual differences among adult women in sensitivity to infants: Evidence in support of an empathy concept. *Journal of Personality and Social Psychology, 46,* 118–124.

Wispé, L. (1986). The distinction between sympathy and empathy: To call forth a concept, a word is needed. *Journal of Personality and Social Psychology, 50,* 314–321.

Received October 26, 1990
Revision received March 10, 1992
Accepted March 30, 1992 ■