

Bryn Mawr College
Scholarship, Research, and Creative Work at Bryn Mawr College

Psychology Faculty Research and Scholarship

Psychology

2004

Reading Others' Emotions: The Role of Intuitive Judgments in Predicting Marital Satisfaction, Quality and Stability

Robert J. Waldinger

Marc S. Schulz
mschulz@brynmawr.edu

Stuart T. Hauser

Joseph P. Allen

Judith A. Crowell

[Let us know how access to this document benefits you.](#)

Follow this and additional works at: http://repository.brynmawr.edu/psych_pubs

 Part of the [Psychology Commons](#)

Custom Citation

Waldinger, R.J., Schulz, M.S., Hauser, S.T., Allen, J.P., & Crowell, J.A. (2004). Reading others' emotions: The role of intuitive judgments in predicting marital satisfaction, quality and stability. *Journal of Family Psychology* 18, 58-71.

This paper is posted at Scholarship, Research, and Creative Work at Bryn Mawr College. http://repository.brynmawr.edu/psych_pubs/18

For more information, please contact repository@brynmawr.edu.

Waldinger, R.J., Schulz, M.S., Hauser, S.T., Allen, J.P., & Crowell, J.A. (2004). Reading others' emotions: The role of intuitive judgments in predicting marital satisfaction, quality and stability. *Journal of Family Psychology, 18*, 58-71.

Running Head: READING OTHERS' EMOTIONS

Reading Others' Emotions: The Role of Intuitive Judgments in
Predicting Marital Satisfaction, Quality and Stability

Robert J. Waldinger

Judge Baker Children's Center

Marc S. Schulz

Bryn Mawr College

Stuart T. Hauser

Judge Baker Children's Center

Joseph P. Allen

University of Virginia

and

Judith A. Crowell

State University of New York at Stony Brook

Abstract

This study examined links between emotion expression in couple interactions and marital quality and stability. Core aspects of emotion expression in marital interactions were identified using naïve observational coding by multiple raters. Judges rated 47 marital discussions using 15 emotion descriptors. Coders' pooled ratings yielded good reliability on four types of emotion expression: hostility, distress, empathy, and affection. These four types were linked with concurrent marital satisfaction and interviewer ratings of marital adjustment, as well as with marital stability at 5-year follow-up. The study also examined the extent to which naïve judges' ratings of emotion expression correspond to "expert" ratings using the Specific Affect Coding System (SPAFF). The unique advantages of naïve coding of emotion expression in marital interaction are discussed.

The U.S. Census Bureau predicts that more than 90% of adults now living in the United States will marry at some point in their lifetimes, and that nearly half of these marriages will end in divorce (Kreider & Fields, 2001). Many of the marriages that remain intact will be characterized by spousal dissatisfaction and poor functioning. Identifying which marriages are likely to succeed and which are likely to fail is an essential component of efforts to prevent marital distress and enhance relationship quality. In observational research on couples, emotional expression has emerged as an important predictor of both marital satisfaction and stability (Gottman, 1994; Smith, Vivian, & O'Leary, 1990). In fact, research suggests that emotional elements of communication may be more highly related to marital quality than actual verbal content (Gottman, 1979; Gottman, 1994; Vivian & O'Leary, 1987), and a recent study suggests that variables derived from observational coding of emotions can predict which couple relationships will remain intact and which dissolve with more than 80% accuracy (Gottman, Coan, Carrere, & Swanson, 1998).

In spite of these findings, researchers have often been reluctant to engage in observational coding of emotional expression in marital interactions, and for good reason. Emotions are evanescent and complicated phenomena. Gathering the kind of data on emotional expression that proves useful in predicting real world outcomes requires a substantial commitment of resources (Fincham, 1998) and poses significant methodological challenges. These challenges include difficult decisions about which elements of emotion to code and how to maximize coders' natural abilities to read emotion while still deriving reliable ratings. The primary aim of the research presented here was to replicate and extend previous studies in which emotional expression was found to be a powerful predictor of both marital distress and dissolution. Our approach differs from traditional studies in that we rely on the consensual judgments of unschooled raters rather

than the “expert” judgments of trained coders. We also sought to add another empirical perspective to an ongoing debate in the marital literature – whether there are particular groupings of emotions beyond positivity and negativity that have implications for marital satisfaction and stability. Finally, we explored the question of how closely a commonly used set of manualized rules for identifying specific emotions corresponds to people's intuitive identification of emotions expressed in interactions.

Our alternative approach to coding using unschooled (naïve) raters is a strategy that has been effectively applied in other areas of psychology to capture elements of non-verbal behavior but has been underutilized in research on emotion in family relationships. This approach takes full advantage of human beings' highly developed natural capacities for instantaneous recognition of emotions.

Emotions and Marital Quality and Stability

Research over the past two decades (e.g., Bradbury & Fincham, 1987; Fincham & Beach, 1999; Jacobson et al., 1994; Thomas, Fletcher, & Lange, 1997) provides clear evidence that emotion is an essential factor to consider in accounting for variability in marital quality (Bradbury, Fincham, & Beach, 2000). However, the specific nature of the association remains uncertain. Positive and negative emotion are intuitively associated with greater and lesser marital quality respectively, and many studies have provided empirical support for these associations (e.g., Holtzworth-Munroe, Stuart, Sandin, Smutzler, & McLaughlin, 1997; Jacob, 1975). However, some research suggests that the categories of positive and negative emotion are too broad to be of maximal predictive utility, and that specific negative emotions (e.g., contempt) and positive emotions (e.g., humor) are critical in predicting marital outcomes (De Koning & Weiss, 1997; Gottman, 1998).

One current controversy in this area concerns the role of expressed anger in eroding or strengthening marriages. Empirical studies have provided inconsistent findings regarding the effect of anger on marital quality and change (e.g., Birchler, Clopton, & Adams, 1984; Notarius, Benson, Sloane, Vanzetti, & Hornyak, 1989). In fact, Gottman and Krokoff (1989) found that anger was related to lower concurrent marital satisfaction but to improvement in marital satisfaction over time, suggesting the possibility that there are different short-term and long-term implications of expressing anger in marital interaction. The work of Gottman and his colleagues has been at the center of the debate about anger. In two studies, they distinguished anger from potentially related negative emotions such as criticism and contempt (Gottman et al., 1998; Gottman, 1994). They found that these latter emotions were reliable predictors of marital dissolution, but that anger was not. They characterized the model explaining these findings as the “specificity of negativity hypothesis” (Gottman, 1998). As several marital researchers have emphasized, (Bradbury et al., 2000; Markman & Notarius, 1987) inconsistent findings regarding the role of emotion in marriage may be related to differences in how emotions are conceptualized and operationalized in various studies.

A brief look at how Gottman operationalizes anger may help clarify important issues in measuring emotion expression. Gottman's *Specific Affects Coding System* (SPAFF, Gottman, McCoy, Coan, & Collier, 1996) is the most fully developed manualized coding system for observing specific emotions. The 169-page SPAFF manual sets forth rules for coding verbal and nonverbal information in order to identify 16 discrete variables. These include emotions such as anger and joy, along with a limited number of behaviors such as criticism and validation that have strong affective implications in marriage and are therefore commonly included in studies of emotions in couple interactions. The SPAFF has separate categories for 10 negative and 5

positive emotions and related behaviors. SPAFF coders rate videotaped marital interactions continuously, judging which of the 16 *mutually exclusive* variables (including “neutral”) is present for each instant of tape.

The SPAFF system distinguishes between five specific negative aggressive emotions or behaviors: anger, contempt, disgust, belligerence, and domineering. Coders are instructed that if anger occurs in conjunction with any other negative code, the other negative code takes precedence (p. 135). Thus, ‘anger’ in the SPAFF system is coded when a moment of interaction does not include contempt, belligerence, or several other highly negative emotions. When coders have parsed negativity into these many forms, what remains in the domain of the SPAFF anger code is likely to be of relatively low intensity.¹ Because the SPAFF system does not code intensity, it is possible that raters may end up distinguishing what most people would refer to as relatively intense anger from relatively benign anger by labeling them as qualitatively different (e.g., by coding more intense anger as “belligerence” and coding less intense anger as “anger”).

Ideally, a coding system for emotional expression would make meaningful qualitative distinctions among specific emotions and capture variations in the intensity of emotion expressed. Capturing variations in emotional intensity is of particular use in understanding emotion regulation. Efforts to modulate intensity are key elements of emotion regulatory processes and are increasingly the focus of emotion (Schulz & Lazarus, in press) and couple interaction research (Burman, Margolin, & John, 1993; Fincham & Beach, 1999). The challenge of developing a reliable coding manual for a system that assesses both specific emotions and a full range of emotional intensity is great, for such a manual would need to include explicit criteria for each emotion as well as specific anchors for the varying intensities of each emotion. This problem and questions about the distinctions researchers have been making among

particular emotions in marital research led us to explore using the pooled judgments of multiple naïve raters, an alternative that has the potential to add important information to our understanding of emotion expression in families.

Using human beings' naturally honed abilities to recognize emotions.

Psychological researchers have used naïve coders for several decades, but rarely to code emotion expression in family interactions (for a notable exception, see Smith et al., 1990). Studies indicate that untutored raters concur remarkably in their judgments of the personality and affective traits of complete strangers. When judgments are pooled, naïve raters exhibit high consensual accuracy and are able to predict important aspects of interpersonal functioning (Albright, Kenny, & Malloy, 1988; Ambady & Rosenthal, 1992; Ambady & Rosenthal, 1993; Paunonen, 1991; Rosenthal, Blanck, & Vannicelli, 1984).

An approach that pools the judgments of untrained raters offers the possibility of maximizing the use of intuitive capacities to judge emotion while minimizing the bias inherent in any one individual's impressions of another's emotions. This method also surmounts the difficulty encountered by some researchers in training raters to code emotion reliably. Smith and his colleagues (1990) note that in attempting to train a group of coders to recognize emotions according to manualized criteria, "the implicit theories of affective expression possessed by the coders were too deeply ingrained for us to alter in a reliable fashion" (p. 792). Coding rules attempt to re-socialize raters to define and recognize emotion in new ways, forcing them to inhibit their own intuitive understanding of emotion in order to carry out the coding task. This need to inhibit native abilities may increase the cognitive demands on coders and may account,

in part, for why it frequently takes a long time for people to learn manual-based emotion coding systems.

Harnessing the predictive power of specific emotions in a limited number of core emotion categories.

Many couple and family studies have focused on the broad dichotomy of positive versus negative emotion in analyzing family interactions (Gottman et al., 1998; Karney & Bradbury, 1997; Markman & Notarius, 1987; Mishler & Waxler, 1968). It is easier for observers to agree on whether an expressed emotion is positive or negative than to agree on finer discriminations within these two broad categories (e.g., between anger and contempt, or between happiness and humor). Although research has provided some support for the predictive utility of making finer distinctions among expressed emotions in marital interactions, large numbers of specific emotions are difficult to analyze statistically. Researchers have looked for ways to group specific emotions together in some meaningful clusters beyond the positive and negative super-categories, but, for the most part, these clusters have been created a priori on primarily theoretical rather than empirical grounds (e.g., Gottman, 1994; Pasch & Bradbury, 1998).

Our approach to observational coding enabled us to explore an empirical basis for a middle ground between the use of a positive/negative emotion dichotomy that is easier to code and analyze but may limit predictive utility and conceptual understanding, and the use of multiple emotion variables that allow for more meaningful predictions about marriage but are more difficult to code reliably and to use in data analysis. We wanted to determine the extent to which untutored raters, using specific emotion labels commonly cited in the marital literature, discriminated among emotions beyond the basic positive and negative valence. In light of the

questions noted above about whether anger is distinct from other negative aggressive emotions with different implications for marriages, we had a particular interest in examining the extent to which naïve individuals distinguished anger from other negative aggressive emotions.

A review of past theory and research on couple and family interactions suggested particular dimensions that might shape the distinctions that lay coders observed in the emotions expressed in couple interactions and might have particularly strong implications for marital quality and stability. *Dominance* has been emphasized as an important relationship dynamic with strong implications for emotion and for quality of functioning. Citing work in child development, social learning theory, and studies of family psychopathology, Markman and Notarius note, "There has been a clear convergence among family scholars regarding the belief that dominance is a key process in family interaction" (Markman & Notarius, 1987, p. 339). Other researchers studying gender differences in emotion in interpersonal interactions have also emphasized the importance of a dominance dimension in distinguishing among different emotions (e.g., Brody, 1999; Timmers, Fischer, & Manstead, 1998). Anger and contempt are emotions typically identified as dominant, whereas emotions such as sadness and fear are seen as more submissive.

In the realm of positive emotion, studies of close relationships have highlighted a potentially important distinction between the expression of *empathy* or validation; and the expression of *affection* or warmth (Linehan, 1987). Whereas affection refers to a feeling of fondness or tender attachment, empathy involves perceiving the internal frame of reference of another with accuracy (Rogers, 1975). Empathy and validation have been defined differently by different investigators, but the core aspect of both terms involves an understanding and recognition of a partner's thoughts and feelings (Levenson & Ruef, 1992). It is possible to feel

affection for a partner without understanding his or her point of view, and the distinction is potentially of great importance in couple relationships. Empathy or validation has been cited by numerous investigators as a key aspect of couples interaction and a predictor of marital satisfaction and functioning (Julien, 1989; Markman & Notarius, 1987; Schaap, 1982; Weiss & Heyman, 1990).

The present study: An alternative approach to examining links between emotional expression and marital satisfaction and stability.

We used the pooled judgments of naïve coders to examine links between expressed emotion and three independently measured marital outcomes: interviewer based assessments of marital quality, self-reports of marital satisfaction, and a five-year follow-up of marital stability. As in manualized approaches to emotion coding, our coders used a set of emotion labels found in previous marital research to be relevant to marital outcomes. However, in contrast to manualized approaches, coders underwent virtually no training but instead depended on their intuitive understandings of commonly used terms (e.g., anger, sadness) to recognize emotions. By pooling the ratings of 5 to 6 coders we were able to derive reliable estimates of the type and intensity of emotional expression in 30-second epochs of interaction.

To ensure a wide range of marital functioning, we chose a sample of young adults that had distinct differences in individual functioning in their adolescence and were now in committed relationships. We also report on a direct comparison of our naïve coding approach with the manualized SPAFF on a small sample of couple interactions that were generously provided by Gottman and Carrere from their studies at the University of Washington. Direct comparison of two observational coding systems across laboratories is rare. Obstacles to such

comparisons include the large time commitment involved in emotion coding and the need to obtain participants' informed consent to use their data in more than one laboratory. In this case, we were fortunate that four couples whose marital interactions were rated by expert SPAFF coders at the University of Washington had given permission for other researchers to study their videotapes.

Method

Participants

Forty-seven heterosexual couples participated in the study. One member of each couple was one of 146 original participants in the Adolescent and Family Development Project (now the Across Generations Project), a longitudinal study of psychological development begun in 1978 (see details in Hauser, with Powers, & Noam, 1991). On entering this longitudinal study at age 14, participants were members of primarily Caucasian middle- and upper-middle-class families. Approximately half were recruited from the freshman class of a local high school ($n=76$), and half were non-psychotic, non-retarded, psychiatrically hospitalized adolescents ($n=70$). The predominant diagnoses during these participants' hospitalizations were mood or disruptive behavior disorders.

In the current study, we examined data from the first 47 original subjects who participated with their romantic partners in follow-up assessments of all participants conducted at age 32. The composition of the sample of 47 original participants was as follows: 20 men and 27 women, 29 from the high school cohort and 18 from the psychiatric cohort. The 94 individuals in these 47 couples were predominantly Caucasian (94%). The average age of participants was 32.2 years ($SD = 3.56$). Thirty-seven of the 47 couples were married (average length of relationship = 5.4 years, $SD = 3.4$ years), while the remaining 10 were living together (average length of relationship = 3.5

years, SD = 3.0 years).² The average number of children per couple was 1.5 (range = 0-5).

Among the 94 participants, the median level of education attained was some years of college without completion of a degree. The median family income was between \$40,000 and \$60,000 per year.

Data from four additional couples were used to compare the SPAFF with our naïve coding system. These couples were participants in a longitudinal study of newlyweds conducted at the University of Washington (Gottman et al., 1998) and had given written permission for other research groups to use videotapes of their interactions for further study. The eight individuals in these couples were mostly Caucasian, middle class (average yearly income between \$40,000 to \$54,000), and college-educated. The mean global score on the Locke-Wallace Marital Adjustment Test (Locke & Wallace, 1959), a widely used measure of marital satisfaction, was 115, which was one standard deviation above the commonly-cited cut point for marital distress (Freeston & Plechaty, 1997).

Procedure

Couple Interaction Task. Participants engaged in two ten-minute laboratory-based discussions of areas of marital conflict, a task widely used in marital research (see Gottman, 1994). Independently of one another, participants were asked to identify the most important areas of disagreement in their current relationship. Participants were asked to discuss the disagreement they rated as most important in a ten-minute videotaped discussion with their partners. Each participant recorded on audiotape a one- or two-sentence statement summarizing the problem to be discussed, and this audiotape was played for the couple at the start of each discussion. In counterbalanced order, couples discussed one problem identified by the man and

one identified by the woman. Among the most common discussion topics were difficulties with couple communication, disagreements over finances, and conflict over household chores.

Discussions took place in a 10 x 12 room in which participants sat facing each other in front of a one-way mirror. Participants were aware that they were being videotaped. Two video cameras were used to obtain clear, close images of each participant's face and top of the torso to optimize the ability to observe facial expression and body language. The two images were recorded in a split-screen format so that partners appeared side by side. Marital discussions in the newlywed study at the University of Washington were conducted according to a similar procedure, but participants discussed one mutually agreed upon topic for 15 minutes (Gottman et al., 1998).

Emotion Coding. The first of the two conflict discussions for each of the 47 couples was rated by undergraduates or recent college graduates, all of whom had completed general course work in psychology. Coders rated participants' emotion expression during the discussion using the emotion expression scales described in the Measures section below. Videotapes of the 10-minute discussions were divided into twenty 30-second segments, and coders rated these segments in randomized order. We chose 30-second segments (as opposed to longer or shorter segments), taking into consideration the amount of time necessary to form an accurate judgment of the emotion being displayed and the practical constraints of the time required to code each segment. Clips were coded in randomized order in an effort to increase the likelihood that sequential connections between segments would reflect ongoing streams of behavior, rather than artifactual connections due to common repeated-measure problems such as carry-over or practice effects. Raters watched each 30-second segment twice, coding first one spouse's emotional expression and then the other. The order was carefully counterbalanced for each segment. To minimize the influence of one partner's behavior on the coding of the other partner, one-half of

the split video screen was covered by dark fabric so that only one participant was visible at a time, but no effort was made to block out the partner's vocalizations.

Two groups of coders produced the data for this study. The first group, consisting of three men and three women, coded the first 40 marital interaction videotapes. A second group of coders was assembled that included two of the original coders along with three new coders. The second group of coders (consisting of two men and three women) rated nine videotapes chosen at random from the original 40, in addition to seven new tapes of couples who participated in the ongoing AGP age 32 assessments after the original group of 40 had been coded.

For the SPAFF comparison, the emotion coding procedures described above were applied to the four marital interaction videotapes provided by the Gottman laboratory. The same videotapes had been coded independently by two experienced SPAFF coders at the University of Washington. The two SPAFF coders rated the videotaped discussions continuously, classifying the participant in each moment of the discussion as expressing one of 16 categories of emotion or emotion-based behavior. Inter-rater reliability for all SPAFF codes as calculated using Cohen's kappa (Cohen, 1968) was .79.³ For this study, data from only one of the two SPAFF coders were used for each participant, and the selection of which coder's data was used was done by random alternation across the 8 participants.

Measures

Emotional expression. The untrained coders were asked to rate participants on 18 variables for each of the 20 segments based on their own understanding of each variable. These 18 variables (see Table 1) were culled primarily from the Specific Affect Coding System (SPAFF, Gottman et al, 1996). Five additional dyadic behavior patterns (e.g., reciprocates

Field Code Changed

partner's negativity) were coded but not analyzed in the study presented. To capture emotional intensity, coders were asked to rate the intensity of the participant's display of each of the 18 variables during that segment of the interaction using Likert-type scales from 0 to 9, with 0 being "not at all" and 9 being "extremely." Coders rated each of these 18 variables separately so that expression of multiple emotions during the 30-second period could be captured easily. No definitions of the individual variables or any additional instructions were given.

Reliability was assessed for each emotion variable using the procedure described by Rosenthal and Rosnow (1991) for calculating the reliability of *composite* scores from multiple raters. Just as the composite reliability of a multi-item scale can be calculated by applying the Spearman-Brown formula to the average inter-*item* correlation and the number of *items* on the scale, the composite reliability of a score aggregated across coders can be determined by using the average inter-rater reliability and the number of *raters*. Pearson correlations were calculated between all possible pairs of coders on each variable for each 30 second-segment of coded videotape. This was initially done using the ratings from the first cohort of coders. Following Rosenthal and Rosnow (1991), the mean inter-rater correlation for each variable was calculated, and the Spearman-Brown formula was applied to these mean correlations to derive a measure of the reliability of the composite scores for each of the 18 coded variables. So, for example, the mean intercorrelation among all pairs of the 6 coders for their ratings of the variable "critical" was .46. Using this correlation and the number of coders (6) in the Spearman-Brown formula yielded a composite reliability of .82 for the pooled ratings of all 6 coders for "critical."

The mean intercorrelation between all possible pairs among the first cohort of six coders for all the variables was .30. Because our final variables combined the ratings from all 6 coders, the mean *composite reliability* of the scores for all the variables was .66, with individual

variables ranging from .89 (humorous) to .27 (disgust) (see Table 1). The mean correlation between all pairs of coders and the composite reliability on the 18 variables were highly similar for the second cohort of coders (mean interrater correlation = .32, composite reliability = .70).⁴

Although our goal was to arrive at reliably coded groupings of emotion that aggregated the individual variables into meaningful clusters (as described below), it is interesting to note that 14 of the 18 individual composite variables had reliabilities of .60 or greater. Due to poor inter-rater reliability, “disgusted” and “belligerent” were dropped from further analyses. Because of their conceptual importance in interpersonal interaction, “fearful” and “tense/anxious,” which were also below generally acceptable levels of reliability, were combined into one anxious/fearful variable, resulting in an effective reliability of .51. This combined variable was included with the other 14 reliable variables in further analyses, resulting in a total of 15 variables.

Relationship Satisfaction. The *Dyadic Adjustment Scale* (DAS) (Spanier, 1976) was used as a measure of marital satisfaction. The DAS is a widely used, 32-item measure of marital satisfaction. It has demonstrated high internal consistency, and has been shown to distinguish between distressed and non-distressed couples, and between abusive and discordant, non-abusive couples (Rosenbaum & O'Leary, 1981). DAS scores range from 0 to 151, with scores below 100 typically used to identify marital distress. DAS data were available on 82 of the 94 participants (41 men and 41 women).⁵ Husbands' and wives' DAS scores were highly correlated ($r = .70$), so scores were averaged to arrive at a couple score for use in all analyses.

Marital Adjustment. The marital adjustment sub-scale of the *Social Adjustment Scale* (SAS) was used as a measure of marital functioning. The SAS (Weissman & Paykel, 1974) is a semi-structured interview that characterizes individuals' adaptive functioning in six domains (e.g.,

work, extended family) on a 7-point scale ranging from excellent (1) to severe impairment (7). Adjustment in the marital domain is assessed by considering the participant's responses to five questions about level of conflict, conflict resolution, and the degree to which the individual's opinions and priorities have been voiced and considered in the relationship in the preceding two months. The global 7-point adaptive functioning rating is based on the *interviewer's* overall assessment of functioning as indicated by responses to these five questions. Lower scores indicate better adjustment. Interviewers were extensively trained using standard procedures (Weissman & Paykel, 1974). In addition, during training, all interviewers independently scored five audiotapes that were part of a separate study (Crowell, Waters, Treboux, O'Connor, & Feider, 1996) and scored by an expert SAS rater. All interviews were audiotaped, and 25 of the tapes were scored by all SAS interviewers to establish interrater reliability. Pearson correlations for all pairs of interviewers were calculated. The mean correlation for ratings of global marital functioning was .86, indicating good inter-rater reliability. Independent interviewers' SAS ratings of husbands and wives were highly correlated ($r = .80$), so scores were averaged to arrive at a couple score for use in all analyses. The marital adjustment score from the SAS was available for 91 of the 94 participants (45 men and 46 women). As expected, marital satisfaction (DAS) and marital adjustment (SAS) scores in this sample were significantly correlated, $r(80) = -.56, p < .01$.

Marital Stability. Participants from the 47 couples included in our sample were contacted by telephone on average 4.7 years ($SD = 1.4$ years) after their laboratory visit as part of the follow-up procedures of the Across Generations Project. At this follow-up, participants were asked whether they were still married or living together with their partner. Thirty-nine of these couples were still together, and eight had separated.⁶

Results

Intensity of Observed Emotional Expression

The first column of Table 1 shows the mean intensities for each of the 18 coded variables averaged over the 20 epochs for each 10-minute discussion. Mean intensities for individual variables were low, indicating minimal expression of any specific emotion or emotion-related behavior in any given 30-second epoch. The generally low levels of expression resulted in some variables having distributions that were positively skewed. For this reason, we conducted a power transformation to approximate normality for the purpose of improving the accuracy of p values and significance levels of tests in our statistical models. All variables were transformed using the formula $\sqrt[2]{x^{2/3}}$ (Box & Cox, 1964).⁷

Identifying core groupings of emotion

Using data from 94 individuals (47 couples), factor analysis was conducted to identify meaningful clusters of the 15 variables. For the factor analysis, participants' mean composite scores over the entire interaction on each of the 15 variables were subjected to principal axis factoring using orthogonal rotation according to varimax criterion.⁸ Based on examination of the Scree plot and using the criterion of Eigen values >1 , four emotion groupings were identified that accounted for 82% of the total variance. The lowest factor loading for any variable was .59, all above the “good” loading level identified by Comrey and Lee (1992). One variable (warm) loaded similarly on two factors, and for theoretical reasons this was included in the factor interpreted as “Affection.” All other variables loaded strongly on only one factor. Table 2 shows the factor loadings for all 15 variables.

Factor 1, which we labeled "Hostility," included the following variables: defensive, critical, angry, irritable, contemptuous, and domineering. Factor 2, labeled "Empathy," included the following variables: acknowledges partner's perspective, interested in understanding partner, and tuned in to partner's feelings. Factor 3 included affectionate, humorous, and warm; and we labeled this factor "Affection." Factor 4 included the following variables: sad, withdrawn, and anxious/fearful. We labeled this factor "Distress." Individual scale scores for each participant were derived by taking the mean of all items on that factor (Tabachnick & Fidell, 1996). Analyses below use these scale scores aggregated over the whole discussion for each participant.⁹ The effective inter-rater reliabilities for these scale scores were all at or above .80 except for Distress, which was .74 (See Table 3). Correlations among the four scales ranged from .11 (Hostility and Distress) to .56 (Affection and Empathy), as shown in Table 3.¹⁰

Anger and other negative aggressive emotions

Given the interest in the field in the relations between anger and other negative aggressive emotions, we examined these linkages more closely. Specifically, we looked at the correlations between anger and criticism, contempt and defensiveness – three of the negative emotions that Gottman distinguishes from anger in his "specificity of negativity" hypothesis. Correlations between anger and these negative aggressive emotions were all large in magnitude for both men and women, sharing as much as 74% of their variance, $r(47) = .86, p < .001$. The only correlation below $r = .74$ was the link between defensiveness and anger for women, $r(47) = .49, p < .001$. These results suggest that unschooled observers do not make clear distinctions between anger and these other negative emotions.

Links Between Emotional Expression and Marital Satisfaction and Functioning. Men in more maritally satisfied couples were seen by coders as expressing greater Empathy, $r(44) = .40$, $p < .01$; greater Distress, $r(44) = .30$, $p < .05$; and less Hostility, $r(44) = -.53$, $p < .001$, in their marital interactions. There was a marginally significant positive association between couple satisfaction and men's expression of Affection, $r(44) = .28$, $p < .07$. Women in more maritally satisfied couples expressed greater Empathy, $r(44) = .35$, $p < .05$, in their marital interactions. There was a marginally significant positive association between couple satisfaction and women's expression of Affection, $r(44) = .25$, $p < .10$. Correlations between couple satisfaction and women's expression of Hostility, $r(44) = -.19$, $p = .22$, and Distress, $r(44) = -.09$, $p = .59$, did not reach the level of statistical significance.

Men who were in couples rated by SAS interviewers as having poorer marital adjustment were observed to express more Hostility, $r(47) = .55$, $p < .001$, and less Empathy, $r(47) = -.43$, $p < .01$, during marital interactions. Women in couples rated by SAS interviewers as having poorer marital adjustment were observed to express more Distress, $r(47) = .50$, $p < .001$, and less Empathy, $r(47) = -.32$, $p < .05$, during marital interactions. Women in less well-adjusted couples, as rated by our interviewers, were marginally less likely to express Affection, $r(47) = -.26$, $p < .10$. The correlation between poorer marital adjustment and the expression of Hostility for women, $r(47) = .14$, $p = .34$, was not statistically significant, nor was the correlation between poorer marital adjustment and the expression of Distress for men, $r(47) = .05$, $p = .73$.

Predicting Marital Dissolution. Couple breakup at follow-up was used as a dichotomous variable, and correlated with each of the four emotion groupings using point biserial correlations. For men, expression of Empathy during the marital interaction was negatively correlated with subsequent break-up, $r(47) = -.29$, $p < .05$, and expression of Affection was negatively correlated

with marital dissolution at a trend level, $r(47) = -.27, p < .10$. Expression of Hostility, $r(47) = -.18, p = .22$, and Distress, $r(47) = -.14, p = .37$, were not significantly correlated with break-up for men. For women, expression of Affection during the marital interaction was significantly negatively correlated with subsequent marital dissolution, $r(47) = -.29, p = .05$; but expression of Hostility, $r(47) = -.01, p = .93$, Distress, $r(47) = .012, p = .93$, and Empathy, $r(47) = -.16, p = .26$ were not significantly linked with break-up.

An additional logistic regression analysis was run to examine the overall predictive power of all four emotion variables in combination.¹¹ Separate models were estimated for women and for men using the four emotion scale scores indexing their emotion expression. In both the men's and the women's models, we correctly identified 84.6% of the couples whose relationships remained intact. Seventy-five percent of the couples whose relationships dissolved were correctly identified using the women's emotion expression variables, while 62.5% were correctly identified using the men's emotion expression variables. Overall, this means that we correctly identified whether couples would remain together or break up by the five-year follow-up with 83% accuracy using women's emotional expression and with 81% accuracy using men's emotional expression. These accuracy rates would be even higher if we combined men's and women's data on emotional expression into one model.

Links Between the Naïve Coding System and the SPAFF

Having provided evidence for the validity of this coding method by using indices of marital functioning, we now turn to the question of how the SPAFF and the naïve coding method compare in their abilities to detect a range of emotional expression. We were particularly

interested in whether the naïve coding system captured low levels of emotion that were not captured by the SPAFF.

We compared the frequency of observed emotions as coded by the SPAFF to the intensity of observed emotions as captured by our naïve coding system using two strategies. First we examined the mean frequencies and intensities of the emotions observed by coders using the two systems. Then we correlated variables derived from both systems to investigate the degree to which the two systems capture similar constructs.

In order to compare SPAFF data with naïve emotion coding data, the 16 SPAFF variables were grouped into scales by summing across items that matched the four factors identified by the principal axis factoring of naïve coding data reported above. *Hostility* was composed of the aggregate frequency of anger, disgust, contempt, belligerence, domineering and defensiveness. *Distress* included sadness, tension, whining and stonewalling. *Empathy* incorporated interest and validation while *Affection* included affection, humor and joy. SPAFF scores for the frequency of expression of Hostility, Distress, Empathy and Affection for each participant were calculated for each 30-second epoch that comprised the coding units for naïve coders. These SPAFF *frequency* scores were compared with naïve coding *intensity* scores.

Table 4 contains descriptive information on the SPAFF-derived frequencies of emotional expression and the intensities of emotional expression derived from naïve coding. In the SPAFF-derived data, hostile and distressing emotions or emotion behaviors were the most frequently coded in each 30-second epoch, although they were coded at low frequency. On average, 2.32 seconds of each 30-second epoch were coded as Hostile and 1.42 seconds were coded as Distress. Empathy and, especially, Affection were coded much less frequently. The generally low mean frequencies for all four categories of emotional expression indicate that much of the

interaction did not contain sufficient emotional expression to trigger a specific affect code by the SPAFF and therefore was coded as neutral. Of the 240 thirty-second epochs, 76.7% were coded as having some kind of emotional expression for at least one second. However, 39% of the epochs in which emotion was identified had frequencies of 1 (21.7%) or 2 (17.4%), indicating that emotion had been coded for only 1 or 2 seconds of that 30 second interval and that neutral was coded for the rest of the epoch. When our four specific categories of emotion are examined, the low frequency of emotion coded becomes even more evident. For example, 97.5% (all but 6) of the 240 thirty-second epochs were seen by SPAFF coders as displaying no Affection and 81.2% were seen as devoid of Empathy.

The average intensities of emotion expressed (see Table 4), as coded by our naïve coding system, were comparable to those obtained for the group of 47 couples. Empathy was the scale rated as being expressed most intensely during the 30-second epochs. Hostility, Distress and Affection were rated as being expressed at significantly lower intensities. The naïve coding system was designed to capture a range of emotion intensities, and it appeared to do this well. For example, all epochs but one were coded as having displayed some Empathy, with 50% of the epochs being coded with an intensity of 2.67 or higher. All but 10 of the epochs (95.8%) were rated as showing evidence of at least some Hostility, and 50% received a score above 0.54. Of the 240 coded epochs, 210 were coded as displaying some Distress and some Affection. Although the distributions yielded by the naïve ratings are still somewhat skewed, there is a substantial range of intensities across epochs, suggesting that differences in intensities between epochs can be meaningfully investigated.

Of the 56 segments coded as neutral on the SPAFF (23.3% of the 240 segments), all were coded as displaying some degree of intensity of positive and negative emotional expression by

our naïve coders. The mean intensities during these SPAFF neutral segments were: Hostility = 0.69; Distress = 0.51; Affection = 0.80; and Empathy = 3.21. This suggests that naïve coding of emotional intensity may meaningfully differentiate low-levels of affective expression that are marked as neutral by the SPAFF system because they fail to exceed the threshold required by the SPAFF system for coding the presence of emotion. It is also possible that this naïve coding system over-identifies emotion, but the meaningful associations with marital quality reported above would argue against this possibility.

The first column of Table 5 contains correlations between the SPAFF and naïve coding emotion expression scores for data at the 30-second epoch level. In these analyses, we correlated emotion scale scores derived from the SPAFF and from the naïve coding system for the 240 epochs coded from the four videotapes of 8 participants. There was a moderate degree of consistency between the frequency with which Hostility was coded using the SPAFF and the intensity of Hostility observed by naïve coders. Similarly, the SPAFF-derived frequency of Empathy also correlated at a moderate level with the intensity of Empathy derived from naïve coding. The connection between the two systems was weak for Distress and absent for Affection. The relatively high number of epochs in the SPAFF data in which no emotion was coded limits the degree of association that can be found at the epoch-level between the SPAFF and naïve coding methods.

As shown in the second column of Table 5, we then examined correlations between the overall frequency and average intensity of Hostility, Distress, Empathy and Affection expressed by each participant over the entire course of the marital discussion. For this analysis, the emotion codes were aggregated across epochs for each individual. Hostility, Distress and Empathy correlated at very high levels, suggesting that the SPAFF and naïve coding methods rank

individuals in highly similar orders in terms of their degree of expression on these three categories of emotion. Ratings of Affection from both systems were moderately correlated even though Affection was coded infrequently using the SPAFF system.

Discussion

In this study we addressed three primary questions: 1) Can we confirm previous findings that emotion expression predicts marital quality and stability? 2) Is it possible to identify groupings of emotions beyond positivity and negativity that are theoretically meaningful and have implications for marital quality and stability? 3) Given the multiplicity of perspectives on emotion, how do the ratings of unschooled coders compare to the ratings of expert coders using manualized rules? In order to address these questions, we developed a method for using the pooled judgments of multiple untrained raters to assess both the intensity and type of emotion expressed in couple interactions.

The predictive power of emotion expression

Consistent with previous research, we found that emotion expressed in marital interactions related in meaningful ways to (1) self-reported marital satisfaction, (2) interview-based assessments of marital quality, and (3) marital dissolution. We found that current marital quality (as measured by the Locke-Wallace and the SAS) was linked to the intensity of expression of four types of emotion. Eight of the 16 correlational links examined (4 emotion scales x 2 indices of marital quality, all calculated separately for men and women) were significant, and three additional correlations were marginally significant. All but one correlation

were in the expected direction, and there was impressive convergence in the pattern of findings across the interview-based and self-report measures.

The finding that men's but not women's Hostility and women's but not men's Distress were significantly correlated with interviewer ratings of poorer marital adjustment may reflect differences in the kinds of emotions that men and women express in distressed marriages.¹² Prior research suggests that at least in some contexts, women may be more likely to express sadness and vulnerability, and men may be more likely to express hostile emotions (Brody, 1999). The unexpected correlation between men's Distress and greater couple satisfaction may be an indicator of men's greater willingness to express vulnerability in more satisfying relationships. In this case, the distinction is between distressing and hostile emotions. By contrast, men's hostile emotions, including anger, were associated with poorer concurrent marital functioning. In fact, men's Hostility accounted for 28% of the variance in the couples' reports of marital satisfaction. These findings lend support to the argument that it is useful to capture distinctions among negative emotions, because not all types of negative emotions function in the same way in marriage.

With regard to our positive emotion groupings, there were connections between both Empathy and Affection and our positive marital outcomes in the expected directions. In addition to the expression of warmth, the expression of the desire to understand one's partner may be particularly helpful to marriages (Gottman, 1994; Julien, 1989; Markman & Notarius, 1987; Schaap, 1982; Weiss & Heyman, 1990). However, the similar pattern of connections and the high degree of correlation between these two groupings suggests some caution in assessing their utility as separate groupings.

Expressed emotions also had implications for marital stability. In fact, the four types of emotion in combination predicted marital dissolution five years hence with more than 80% accuracy. This finding is particularly impressive when one considers that the coders were untrained college-age young adults. Our correlational analyses indicated that the expression of positive emotions was significantly linked with marital stability, whereas the expression of negative emotions was not. For men the expression of greater Empathy (and at the trend level, greater Affection), and for women the expression of greater Affection predicted that the couple would remain together in the five years following the observed interaction. This finding is consistent with those of the Gottman et al. newlywed study (1998) and the Pasch and Bradbury (1998) study of social support in marriage, both of which found that positive emotions and behaviors predicted marital stability. Increasing attention to the role of positive and supportive behaviors in marriage is clearly warranted (Cutrona, 1996).

Finding a middle ground – Hostility, Distress, Affection, and Empathy

Researchers continue to search for an appropriate compromise between the convenience of parsing expressed emotion into two global groupings of positivity and negativity, and the promise of greater understanding that might come with considering multiple specific emotions. The findings reported above indicate why such a compromise may be useful. The ratings of our untrained judges clustered into four emotion groupings that were differentially linked with key indices of marital functioning. Moreover, the four groupings make sense in light of prior theory and research on dimensions of emotion in interpersonal interactions. Our raters clearly distinguished between two types of negative emotion – Hostility and Distress. It may be argued that Hostility and Distress differ most markedly along the dominance dimension found by

Russell and Mehrabian (1974) and discussed extensively in the literature on gender differences in interaction (Brody, 1999; Timmers et al., 1998). Our naïve raters did not clearly distinguish anger from other negative aggressive emotions such as criticism and contempt. This finding suggests caution in differentiating anger from related negative aggressive emotions. Recent findings using the SPAFF have suggested that anger is not “toxic” to marriages but that criticism, contempt, defensiveness, and withdrawal are. It is possible that “anger” as defined by the SPAFF is not qualitatively different from these other negative emotions, but that it simply represents a less intense form of negative aggressive emotionality. Because of the potentially important clinical implications of this issue, more research is warranted.

Our untrained raters also distinguished between two types of positive emotions or emotion relevant behaviors – Affection and Empathy. Although caution is warranted because these groupings of emotion correlated within this sample at a level of .56 and other research has suggested the presence of only one positive factor (Smith et al., 1990), this distinction is supported by other empirical investigations (Gottman, 1994; Julien, 1989; Markman & Notarius, 1987; Schaap, 1982; Weiss & Heyman, 1990) in which empathy (or validation) and affection (or warmth) differ in the extent to which they predict marital stability and satisfaction. Further research is needed in which additional emotion descriptors conceptually linked with empathy or affection are incorporated into the rating system, so that the degree of overlap or independence can be clarified.

Comparing the perspectives of naïve and expert coders

Our naïve coding approach also had predicted links with the SPAFF, despite the fact that the SPAFF system codes for the presence or absence of an emotion and our naïve coding

approach rates intensity. Although comparison of the two systems at the level of 30-second segments of a couple's interaction yielded relatively low correlations, we found high correlations between the two sets of ratings for participants over the entire course of a discussion. This suggests that the two coding methods are consistent in their ratings of the degree to which individuals express particular emotions in a marital interaction. SPAFF data, collected second-by-second, are particularly useful for fine-grained sequential analysis of emotion patterns within dyads. Data from the naïve coding system appear well suited to examining the intensity of emotional expression, an aspect of emotion that is especially relevant in the study of emotion regulation. We recognize of course that the sample was quite small due to practical constraints noted above, and we must consider these results with caution. However, the magnitude of links between the two systems at the participant level was very strong.

The magnitude and consistency of the links between our four emotion groupings and (a) the SPAFF and (b) both current indices of relationship functioning and long-term marital stability provide compelling evidence for the validity of these groupings. Pooling the ratings of untrained coders yields reliable estimates of the intensity with which 14 different specific emotions or emotion relevant behaviors were expressed in marital interactions. It is noteworthy that reliability was calculated at the level of 30-second segments of the discussion and would be higher for ratings pooled for the entire marital discussion. Good reliability of these individual variables at the 30-second epoch level allows for the possibility of using these data to examine sequences of emotional expression in couples – for example, patterns of reciprocation of negative emotions that have been found in other studies to predict marital dissatisfaction and divorce (Gottman, 1994).

We believe that naïve coding has both theoretical and practical advantages for researchers studying emotional expression in couples. The ecological validity of a cultural informants' approach is a particular strength of this method. By not demanding adherence to a prescribed set of rules, naïve coding takes advantage of human beings' well-honed and highly adaptive abilities to read others' expressions of emotion. The emotion groupings that naïve coders consistently identified in this research may represent fundamental typologies of emotional expression that guide people's evaluations of interactions in close relationships. Not coincidentally, the groupings identified – Hostility, Distress, Affection, and Empathy – are directly linked to the descriptors that individuals commonly use to characterize the interpersonal style of others. Research on marital interaction is likely to benefit from consideration of these readily identifiable types of emotional expression.

We see several practical strengths that our naïve coding system brings to the study of emotional expression in the particular context of marital interactions. Coders require little or no training. Good to excellent effective inter-rater reliability can be achieved using composite scores obtained from multiple raters, especially for the larger emotion clusters we have identified. The system allows for incorporation of multiple perspectives within the group of coders, which is particularly important given gender and cultural differences in assessing emotional expression in marriage. The system also is sensitive to low intensities of emotional expression. These low intensity emotional displays may be quite meaningful for the spouses in the interaction and therefore important for researchers to identify. By capturing low intensity and variations in intensity across the range of emotional expression, the naïve coding approach also permits researchers to examine more carefully processes of emotion regulation in couples' relationships. Moreover, in contrast to the SPAFF, our naïve rating system also allows coders to

rate the intensity of multiple emotions at the same time, allowing us to gather data about emotions that are expressed virtually simultaneously ("emotion blends").

Manualized coding systems that employ a binary decision making framework (i.e., coding for the presence or absence of emotion) must include clear guidelines about the level of emotion required to trigger an emotion code. In such systems, there is likely to be subtle pressure to establish a relatively high threshold for triggering an emotion code. Instructing raters to code an emotion only when there is overwhelming evidence for its presence (e.g., telling raters to code an emotion only when it "hits you over the head") may increase inter-rater reliability. This added reliability, however, comes at the cost of losing information about low intensity emotion displays. The naïve coding approach presented in this report, which involves assessments of varying intensities rather than the presence or absence of emotion, appears to yield good levels of inter-rater reliability and to capture meaningful but low levels of emotion expression.

Perhaps the greatest disadvantage of the naïve coding approach is the task of coordinating the efforts of multiple coders. This effort must be weighed against the extensive training and reliability testing required by manualized coding systems. Based on our experience, two alterations in the coding method presented here may be warranted. Although coding segments of a videotaped discussion in random order may help minimize carry-over effects from one segment of tape to another, this advantage may be outweighed by the importance of seeing each segment of a discussion in the context of what has come before it. Similarly, covering half of the video screen so that only one participant is visible for coding may be unnecessary and may hinder assessment of the full context of an individual's emotional expression. Both of these issues will be tested empirically in future investigations.

It is important to keep in mind several limitations of this work. The sample of 47 couples, while representing a range of psychological functioning, is small for the use of factor analytic and logistic regression techniques. Future replication will be important. We also recognize that the choice of individual variables that we offered to naïve coders undoubtedly affected the factor structure that emerged. These variables had been shown in prior couples research to be relevant to marital functioning, but it is possible that the inclusion of other emotion labels would result in different emotion groupings. Ideally, we would have compared the SPAFF-derived emotion variables with our naïve coding approach using a larger sample. However, the hurdles to comparing two microanalytic observational coding systems are so great and the direct comparison of such systems is so rare that we believed these analyses warranted presentation. It is possible that the strength of some of the concurrent and predictive validity findings is due, in part, to our use of two cohorts that originally differed in their levels of functioning in important ways. However, it is important to recognize that these differences were during adolescence, at least 15 years prior to the timing of the assessments used in this study.

Despite these limitations, the results of this study are noteworthy in several respects. Similar to other studies, we found that emotional expression was linked to concurrent marital quality and to relationship stability over a nearly five-year period. In contrast to other studies, our findings are based on the judgments of untrained college-age young adults. That is, the judgments of unmarried college-age individuals, when pooled appropriately, tell us a great deal about how couples are doing and about the likelihood of them remaining together. These pooled judgments also shed light on four types of emotional expression that may be particularly salient in marital interactions and can predict long-term marital stability with considerable accuracy. The relevance of these four groupings of emotion to marital functioning and satisfaction is

consistent with prior research on couples and families. In emphasizing these four emotion groupings, we do not deny the importance of studying discrete emotions. However, most couple and family researchers have, of necessity, reduced data on discrete emotions into larger categories. Many researchers have hesitated to engage in observational coding of emotions because of the myriad emotion variables that could be coded. This study provides empirical support for focusing on four categories of emotional expression in couples' interactions that are characterized by theoretically meaningful distinctions among discrete emotions.

Implications for Application and Public Policy

Interventions to help couples modify emotion expression and regulation processes are key elements of many approaches to marital therapy (Christensen & Jacobson, 2000; Gottman, 1999; Stanley, Blumberg, & Markman, 1999). The findings of this study point to central elements of emotional expression in marital interaction that clinicians may want to consider in their work with distressed couples. This study suggests that it is important to move beyond a simple focus on positive versus negative emotions to a more differentiated perspective that distinguishes between emotions associated with hostility and distress on the negative side, and between affection and empathy on the positive side. Just as the results of this study can guide the efforts of future observational coding of marital interactions, these findings can also help clinicians focus on fundamental aspects of emotion expression that may be important to marital functioning and stability.

References

- Albright, L., Kenny, D., & Malloy, T. (1988). Consensus in personality judgments at zero acquaintance. Journal of Personality and Social Psychology, *55*, 387-395.
- Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequence: A meta-analysis. Psychological Bulletin, *111*(2), 256-274.
- Ambady, N., & Rosenthal, R. (1993). Half a minute: Predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness. Journal of Personality and Social Psychology, *64*, 431-441.
- Birchler, G. R., Clopton, P. L., & Adams, N. L. (1984). Marital conflict resolution: Factors influencing concordance between partners and trained coders. American Journal of Family Therapy, *12*, 15-28.
- Box, G., & Cox, D. (1964). An analysis of transformations. Journal of the Royal Statistical Society, Series B.
- Bradbury, T. N., & Fincham, F. D. (1987). Assessment of affect in marriage. In K. D. O'Leary (Ed.), Assessment of marital discord: An integration for research and clinical practice (pp. 59-108). Hillsdale, NJ: Erlbaum.
- Bradbury, T. N., Fincham, F. D., & Beach, S. R. H. (2000). Research on the nature and determinants of marital satisfaction: A decade in review. Journal of Marriage & the Family, *62*(4), 964-980.
- Brody, L. (1999). Gender, emotion and the family. Cambridge, MA: Harvard University Press.

- Burman, B., Margolin, G., & John, R. S. (1993). America's angriest home videos: Behavioral contingencies observed in home reenactments of marital conflict. Journal of Consulting and Clinical Psychology, *61*(1), 28-39.
- Christensen, A., & Jacobson, N. (2000). Reconcilable differences. New York: Guilford Press.
- Cohen, J. (1968). Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. Psychological Bulletin, *70*(4), 213-220.
- Comrey, A., & Lee, H. (1992). A first course in factor analysis (2nd ed.). Hillsdale, NJ: Erlbaum.
- Crowell, J. A., Waters, E., Treboux, D., O'Connor, E., Colon-Downs, Christina, & Feider, O. (1996). Discriminant validity of the adult attachment interview. Child Development, *67*, 2584-2599.
- Cutrona, C. (1996). Social support in couples. New York: Sage.
- De Koning, E., & Weiss, R. (1997). A funny thing happened during my marriage. Paper presented at the meeting of the Association for the Advancement of Behavior Therapy, Miami, Florida.
- Fincham, F. (1998). Child development and marital relations. Child Development, *69*(2), 543-574.
- Fincham, F. D., & Beach, S. R. H. (1999). Conflict in marriage: Implications for working with couples. Annual Review of Psychology, *50*, 47-77.
- Freeston, M. H., & Plechaty, M. (1997). Reconsideration of the Locke-Wallace marital adjustment test: Is it still relevant for the 1990s? Psychological Reports, *81*, 419-434.

- Gottman, J. (1979). Marital interaction: Experimental investigations. New York: Academic Press.
- Gottman, J. (1998). Psychology and the study of marital processes. Annual Review of Psychology, 49, 169-197.
- Gottman, J. (1999). The marriage clinic: A scientifically based marital therapy. New York: W.W. Norton & Co.
- Gottman, J., Coan, J., Carrere, S., & Swanson, C. (1998). Predicting marital happiness and stability from newlywed interactions. Journal of Marriage & the Family, 60(1), 5-22.
- Gottman, J., McCoy, K., Coan, J., & Collier, H. (1996). The specific affect coding system (spaff). In J. Gottman (Ed.), What predicts divorce? The measures (pp. 1-169). Hillsdale, NJ: Erlbaum.
- Gottman, J. M. (1994). What predicts divorce?: The relationship between marital processes and marital outcomes. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Gottman, J. M., & Krokoff, L. (1989). Marital interaction and marital satisfaction: A longitudinal view. Journal of Consulting and Clinical Psychology, 57, 47-52.
- Hauser, S. T., with Powers, S., & Noam, G. (1991). Adolescents and their families: Paths of ego development. New York: Free Press.
- Holtzworth-Munroe, A., Stuart, G., Sandin, E., Smutzler, N., & McLaughlin, W. (1997). Comparing the social support behaviors of violent and nonviolent husbands during discussions of wife personal problems. Personal Relationships, 4, 395-412.
- Jacob, T. (1975). Family interaction in disturbed and normal families: A methodological and substantive review. Psychological Bulletin, 82, 33-65.

Jacobson, N. S., Gottman, J. M., Waltz, J., Rushe, R., Babcock, J., & Holtzworth-Munroe, A. (1994). Affect, verbal content, and psychophysiology in the arguments of couples with a violent husband. Journal of Consulting and Clinical Psychology, *62*(5), 982-988.

Julien, D. (1989). A comparison of a global and a microanalytic coding system: Implications for future trends in studying interactions. Behavioral Assessment, *11*, 81-100.

Karney, B. R., & Bradbury, T. N. (1997). Neuroticism, marital interaction, and the trajectory of marital satisfaction. Journal of Personality & Social Psychology, *72*, 1075-1092.

Kreider, R., & Fields, J. (2001). Number, timing, and duration of marriages and divorces: Fall 1996 (Current Population Reports, P70-80). Washington, DC: U.S. Census Bureau.

Levenson, R., & Ruef, A. (1992). Empathy: A physiological substrate. Journal of Personality & Social Psychology, *63*, 234-246.

Linehan, M. M. (1987). Dialectical behavior therapy for borderline personality disorder. Bulletin of the Menninger Clinic, *51*, 261-276.

Locke, H., & Wallace, K. (1959). Short marital adjustment and prediction tests: Their reliability and validity. Marriage and Family Living, *2*, 251-255.

Markman, H. J., & Notarius, C. I. (1987). Coding marital and family interaction: Current status. In T. Jacob (Ed.), Family interaction and psychopathology: Theories, methods, and findings. (pp. 329-390): Plenum Publishing.

Mishler, E. G., & Waxler, N. W. (1968). Interaction in families: An experimental study of family processes and schizophrenia. New York: Wiley.

Notarius, C. I., Benson, P. R., Sloane, D., Vanzetti, N. A., & Hornyak, L. M. (1989). Exploring the interface between perception and behavior: An analysis of marital interaction in distressed and nondistressed couples. Behavioral Assessment, *II*, 39-64.

Pasch, L. A., & Bradbury, T. N. (1998). Social support, conflict and the development of marital dysfunction. Journal of Consulting and Clinical Psychology, *66*(2), 219-230.

Paunonen, S. (1991). On the accuracy of ratings of personality by strangers. Journal of Personality and Social Psychology, *61*, 471-477.

Rogers, C. (1975). Empathic: An unappreciated way of being. Counseling Psychologist, *5*, 2-10.

Rosenbaum, A., & O'Leary, K. (1981). Children: The unintended victims of marital violence. American Journal of Orthopsychiatry, *51*, 692-699.

Rosenthal, R., Blanck, P. D., & Vannicelli, M. (1984). Speaking to and about patients: Predicting therapists' tone of voice. Journal of Consulting and Clinical Psychology, *52*, 679-686.

Rosenthal, R., & Rosnow, R. L. (1991). Essentials of behavioral research: Methods and data analysis (2nd ed.). New York: McGraw-Hill.

Russell, J., & Mehrabian, A. (1974). Distinguishing anger and anxiety in terms of emotional response factors. Journal of Consulting and Clinical Psychology, *42*, 79-83.

Schaap, C. (1982). Communication and adjustment in marriage. The Netherlands: Swets and Feitlinger.

Schulz, M., & Lazarus, R. (in press). Emotion regulation during adolescence: A cognitive-mediational conceptualization. In A. Cauce & S. Hauser (Eds.), Adolescence and beyond: Family interactions and transitions to adulthood, advances in family research. Mahwah, NJ: Lawrence Erlbaum Associates.

Smith, D., Vivian, D., & O'Leary, K. D. (1990). Longitudinal prediction of marital discord from premarital expressions of affect. Journal of Consulting and Clinical Psychology, *58*, 790-798.

Spanier, G. (1976). Measuring dyadic adjustment: New scales for assessing the quality of marriage and similar dyads. Journal of Marriage and the Family, 38, 15-28.

Stanley, S., Blumberg, S., & Markman, H. (1999). Helping couples fight for their marriages: The prep approach. In R. Berger & M. Hannah (Eds.), Preventive approaches in couples therapy (pp. 279-303). Philadelphia: Brunner/Mazel.

Tabachnick, B., & Fidell, L. (1996). Using multivariate statistics (3rd ed.). New York: Harper Collins.

Thomas, G., Fletcher, G., & Lange, C. (1997). On-line empathic accuracy in marital interaction. Journal of Personality and Social Psychology, 72(4), 839-850.

Timmers, M., Fischer, A., & Manstead, A. (1998). Gender differences in motives for regulating emotions. Personality and Social Psychology Bulletin, 24, 974-985.

Vivian, D., & O'Leary, K. (1987). Communication patterns in physically aggressive engaged couples. Durham, NH: University of New Hampshire Third National Family Violence Research Conference.

Weiss, R., & Heyman, R. (1990). Observation of marital interaction. In F. Fincham & T. Bradbury (Eds.), The psychology of marriage: Basic issues and applications (pp. 87-117). New York: Guilford.

Weissman, M., & Paykel, E. (1974). The depressed woman: A study of social relationships. Chicago: University of Chicago Press.

Author Note

Robert J. Waldinger, Judge Baker Children's Center and Harvard Medical School; Marc S. Schulz, Department of Psychology, Bryn Mawr College; Stuart T. Hauser, Judge Baker Children's Center and Harvard Medical School; Joseph P. Allen, Department of Psychology, University of Virginia; Judith A. Crowell, Department of Psychiatry, State University of New York at Stony Brook.

The study was supported by grants from the National Institute of Mental Health (K08 MH 01555 and MH 44934-11). We gratefully acknowledge the contributions of the following people whose advice and assistance contributed significantly to this project: Sybil Carrere, John Gottman, Heidi Gralinski-Bakker, Carl Morris, Robert Rosenthal, Katie Swanson, and Erica Woodin.

Correspondence concerning this article should be addressed to Robert J. Waldinger, Judge Baker Children's Center, 3 Blackfan Circle, Boston, Massachusetts 02115, Telephone: (617) 232-8390 ext. 2836, email: rwalding@jbcc.harvard.edu.

Table 1

Means and Reliability of 18 Emotion Variables

Emotion Variable	Intensity of expression				Average correlation between coders	Inter-rater reliability of composite scores
	Score		Transformed ^a score			
	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>		
Defensive	1.27	0.77	2.25	0.91	.37	.76
Critical	1.32	0.96	2.29	1.09	.46	.82
Affectionate	0.47	0.56	1.09	0.78	.41	.79
Angry	0.36	0.54	0.82	0.84	.37	.76
Sad	0.28	0.42	0.71	0.71	.26	.65
Warm	1.02	0.64	1.93	0.84	.34	.74
Tense/anxious	0.87	0.55	1.74	0.77	.15	.47
Irritable	0.41	0.45	0.98	0.73	.25	.65
Humorous	0.65	0.72	1.35	0.96	.60	.89
Acknowledges partner's perspective	2.42	0.95	3.54	0.99	.36	.75
Withdrawn	0.61	0.67	1.27	0.96	.31	.71
Contemptuous	0.31	0.50	0.75	0.77	.32	.75
Interested in understanding partner	2.34	1.07	3.43	1.14	.35	.73
Fearful	0.14	0.22	0.45	0.43	.10	.37
Domineering	0.62	0.76	1.28	1.01	.37	.76
Belligerent	0.11	0.18	0.36	0.41	.16	.49

Reading Others' Emotions 42

Disgusted	0.14	0.22	0.44	0.46	.07	.27
Tuned in to partner's feelings	1.48	0.73	2.51	0.94	.21	.60
Overall Mean					.30	.66

^a Data were transformed using the formula $\frac{2x}{2.3}$.

Table 2

Principal Axis Factor Analysis of Emotion Expression Variables^a (94 Participants)

Emotion Variable	Factor			
	1	2	3	4
Critical	<u>.96</u>	-	-.11	-
Contemptuous	<u>.82</u>	-.31	-	-
Angry	<u>.77</u>	-.19	-.21	.25
Irritable	<u>.77</u>	-.36	-.11	.22
Domineering	<u>.77</u>	-	-	-.36
Defensive	<u>.68</u>	-.17	-	.25
Tuned in to partner's feelings	-.22	<u>.90</u>	.13	-.14
Interest in understanding partner	-.30	<u>.86</u>	-	-.27
Acknowledges partner's perspective	-.34	<u>.86</u>	.12	-.18
Warm	-	.72	<u>.59</u>	-
Humorous	-.20	-	<u>.84</u>	-
Affectionate	-.11	.43	<u>.75</u>	-
Anxious/Fearful	-	-.28	.26	<u>.66</u>
Sad	.20	-	-.25	<u>.64</u>
Withdrawn	-.13	-.28	-.12	<u>.63</u>

Note. The Extraction method used was Principal Axis Factoring. The rotation method used was Varimax with Kaiser Normalization. Italicized values represent factor loadings used to make up the factors in each column. Dashes represent factor loadings $<.10$.

Table 3

Mean Scores, Inter-rater Reliabilities and Pearson Correlations Among Four Emotion Composite Variables (94 Participants)

Emotion Factor	<u>M</u>	<u>SD</u>	Inter-rater reliability of composite scores	Correlations		
				Hostility	Distress	Affection
Hostility	0.79	0.58	.87			
Distress	0.49	0.38	.74	.11		
Affection	0.79	0.59	.87	-.33	-.25	
Empathy	2.22	0.95	.80	-.48	-.38	.56

Note: *P* values are not reported because husbands and wives are not independent of each other and standard *p* values are therefore inappropriate.

Table 4

Means, Standard Deviations, and Percent of Total Epochs with No Emotion for SPAFF and Naïve Coding Systems

Emotion Factor	SPAFF			Naïve Coding		
	Average ^a		% Epochs with no emotion	Average ^b		% Epochs with no emotion
	Frequency			Intensity		
	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>		
Hostility	2.32	4.48	64.6	.97	.94	4.2
Distress	1.42	4.48	62.5	.45	.49	12.5
Affection	.06	.53	97.5	.74	.65	12.5
Empathy	.46	3.31	81.2	2.83	1.28	0.4

Note. Two hundred forty 30-second epochs were gathered from eight participants in four interactions.

^aThe SPAFF-derived scores represent the average frequencies of coded emotions in each emotion category for each 30-second epoch. This frequency score can also be interpreted as the number of seconds within the 30-second interval in which an emotion in that category was coded.

^bThe scores derived from naïve coding are simply the mean intensity score for emotion variables in each emotion category for the 30-second epochs in which the participants' interactions were rated.

These data have been transformed using the formula $\frac{2x}{3}$.

Table 5

Correlations between SPAFF Emotion Frequencies and Naïve Emotion Coding Intensities at 30-Second Epoch and Participant Levels

Emotion Factor	Correlations	
	30-second epochs N=240	Aggregated by participant N=8
Hostility	.44	.95
Distress	.15	.75
Affection	.03	.45
Empathy	.36	.82

Note: *P* values are not reported because the 240 epochs are nested within 4 couples and are therefore not independent. Traditional *p* values are likely to overestimate the true significance of these correlations.

Footnotes

¹ Gottman and his colleagues (1998) appeared to arrive at a similar conclusion when they experimented with giving numerical weights to specific SPAFF variables based on their empirical correlations with marital satisfaction in previous research. They defined high-intensity negativity as contempt, defensiveness, and belligerence whereas anger was classified low intensity negativity.

² Because the majority of couples were married, we will refer to the partners in all relationships as husbands and wives to facilitate fluency of writing.

³ More information on the reliability of this coding can be found in previous reports by Gottman and colleagues (J. Gottman, J. Coan, S. Carrere, & C. Swanson, 1998; Gottman, Swanson, & Murray, 1999).

⁴ The average pairwise correlation between old and new coders on the 18 overlapping participants was .30, which is of a similar magnitude as the average pairwise correlation among the original coders of the first 80 participants, suggesting adequate reliability. However, paired t-tests comparing the original coders averages to the new coders averages on the 15 emotion variables yielded differences on the majority of the variables; the new raters systematically assigned higher scores, suggesting a cohort effect. To equalize the metrics used by old and new coders we calculated the ratio of the mean score for the old raters to the mean score of the new raters on each of the 15 variables. We then transformed all new coders' data by multiplying their scores by the appropriate ratio for each variable. As with any Likert-type scale, it is important to

be cautious about reifying the meaning of the absolute score. We do not, therefore, attach particular significance to the difference in absolute scores given by either cohort. Rather, our goal was to gauge reliably the *variability* in emotion expression across 30-second epochs and across individual participants. What is important is that this variability was measured consistently across coding groups, and our analyses indicate that it was.

⁵ T-tests comparing the emotion scale scores of the 12 participants for whom DAS data were missing with those of the other 82 participants revealed no significant differences.

⁶ An ANOVA revealed no significant link between breaking up and the time between the marital interaction and telephone follow-up.

⁷ Data transformation was based on inspection of the data and was carried out according to procedures recommended in Box & Cox (1964) and Tabachnick & Fidell (1996).

⁸ Principal Axis Factor Analysis was used because of our interest in extracting all meaningful theoretical factors from the data. Additional principal components analyses produced a substantially similar solution.

⁹ Inspection of the distribution of scores on the emotion expression scales revealed two significant outliers. For men, a single outlying score on Distress was four standard deviations above the mean; and for women, a single outlying score on Hostility was 4.5 standard deviations above the mean. In accordance with procedures outlined in Tabachnick and Fidell (1996), we

transformed each score so that it was 2 standard deviations above the mean, and these were used in subsequent analyses.

¹⁰ Consistent with the factor analytic results, the mean intercorrelation among variables within each scale (median $r = .66$) was noticeably higher than the mean intercorrelation among variables from different scales (median $r = .22$). (Median r 's were calculated using the absolute values of r .)

¹¹ Our final models included all significant interactions among the four variables.

¹² The correlations of men's and women's Hostility with marital adjustment were significantly different from each other, $t(43)=7.24, p<.01$, as were the correlations of men's and women's Distress with marital adjustment, $t(43)=-7.37, p<.01$.