RACHEL EBLING AND ROBERT W. LEVENSON
*University of California—Berkeley*

# Who Are the Marital Experts?

*We asked whether professional training or personal experience with marriage predicted accuracy in judging (a) marital satisfaction and (b) marital stability. Nine groups of participants viewed 3-minute videotaped conversations of 10 married couples and rated each on level of marital satisfaction and whether they were likely to divorce in the future. Group differences were found in accuracy of marital satisfaction judgments. Those for whom marriage held high* personal *meaning (satisfied and dissatisfied long-term marriages, newlyweds, recent divorcé[e]s), as rated by a panel of judges, were more accurate than those with* professional *training (pastoral counselors, clinical psychology graduate students, marital therapists, marital researchers). Neither professional training nor personal experience was associated with the ability to predict divorce.*

Over the past four decades, considerable research has been directed toward two qualities of marriage: (a) marital satisfaction—finding characteristics that distinguish satisfied marriages from dissatisfied marriages—and (b) marital stability—finding characteristics that distinguish couples who stay together from those who divorce. Although these two qualities often go hand in hand, they sometimes do not. For example, we all know of marriages in which the spouses are quite unhappy but yet they stay together over the years.

Department of Psychology, University of California, Berkeley, California, 94720 (rebling@socrates.berkeley.edu).

Most of the early research on marital satisfaction and marital stability was done by sociologists who made use of well-constructed self-report instruments to study marriage and predict its course. A major change came in the 1970s, when behavioral observation was introduced into marital research (e.g., Raush, Barry, Hertel, & Swain, 1974). With this change, investigators could examine marital interaction and try to find behavioral signs that were related to marital quality and marital stability. Behavioral observation research, which most often made use of quite laborious and time-consuming coding of marital interactive behavior, identified a number of potent predictors of both marital satisfaction and marital stability that were often superior in predictive power to those derived from self-report measures (see Gottman, 1998, for a review). More recently, researchers in the social-cognition tradition identified qualities of marital cognitions and attributions that were also associated with marital satisfaction and stability (Fincham & Bradbury, 1991).

All of these approaches to predicting marital satisfaction and stability are based on the application of highly refined measurement techniques developed by marital researchers and applied in well-controlled studies. People in everyday social settings make similar informal assessments of couples they encounter—including judgments about whether couples seem happy together and whether relationships are likely to last. These *real world* judgments are usually made on the basis of brief observations of marital interaction without the aid of questionnaires, systematic behavioral coding, or assessment of marital attributions, and they are often stated with considerable authority

and confidence. Yet we rarely get definitive feedback about whether such judgments are accurate. For example, our judgments of couples' marital satisfaction are typically not checked against the couples' own feelings in this regard and, assuming we remembered them, we would not know whether our judgments of marital stability are accurate until many years had passed.

If we grant that laypeople regularly make judgments about marriages, it raises the following question: On what are these judgments based? Several studies have examined the use of actual and perceived behavioral cues in making judgments about married couples. Kleinke, Meeker, and La Fong (1974) found that untrained judges made more positive ratings of couples who in fact gazed at and touched one another than they did for nongazing and nontouching couples. Royce and Weiss (1975) found that untrained judges used couples' actual aversive behaviors and tended to ignore actual supportive behaviors when assessing marital satisfaction. In a variation of this methodology, Wills, Weiss, and Patterson (1974) examined the relationship between marital satisfaction and spouses' daily ratings of their partners' perceived marital behavior. They found that ratings of partners' negative affective behaviors and negative instrumental behaviors were the best predictors of marital dissatisfaction.

## WHO MAKES ACCURATE JUDGMENTS?

This study is based on the premise that when it comes to making accurate real world marital judgments, we are not all equals. Rather, it seems likely that some people, by virtue of their training or life experiences, would be more accurate than others. One obvious group that might qualify for the status of *marital experts* would be those who have professional training in the understanding and changing of human behavior. The validity of clinical judgments has become an increasingly popular topic of research. A good deal of effort has been directed toward examining the relationship between clinical training or experience and accuracy in performing such professional functions as personality assessment, clinical diagnosis, and behavioral prediction. Overall, these research efforts have yielded a mixed picture. For example, Peterson (1995) has argued that there is a substantial empirical basis for the idea that professional training in psychology is unrelated to validity of clinical judgments. Garb's (1989) review of the literature led him to conclude that experienced

clinicians were generally not more accurate in their judgments than less experienced clinicians or graduate students, but that clinicians were sometimes more accurate than laypersons. More specifically, Garb noted that clinically trained judges (e.g., psychiatrists, clinical psychologists, social workers) were more accurate than lay judges (e.g., undergraduates, ministers, secretaries) in only 10 of the 30 studies he reviewed.

Thus whether professional training allows one to make more valid judgments about other people is still an open question. Moreover, studies that have compared various types of professional groups against laypersons have treated laypersons as a homogeneous entity. The possibility that there are differences in accuracy among subgroups of laypersons who differ in experience relevant to the judgment task (e.g., spouses in long-term marriages in a judgment task about marriage) has not been investigated.

When it comes to the nature of the judgments themselves, the vast majority of formal studies of accuracy have concerned judgments about qualities of individuals, rather than qualities of married couples. Among investigations of the latter type, Royce and Weiss (1975) provided one of the few studies that assessed the accuracy of ratings of marital satisfaction, finding that untrained laypeople could identify distressed and nondistressed couples with an accuracy rate of 63% (vs. a chance level of 50%). Other studies of marital judgments have not assessed the accuracy of judgments of marital satisfaction, but rather the accuracy of ratings of marital behavior. For example, Floyd and Markman (1983) found that spouses' evaluations of their own behavior were more accurate than evaluations of their partners' behavior. Margolin, Hattem, John, and Yost (1985) found that spouses' observations of marital behavior were more similar to that of trained coders when judging a stranger couple, rather than themselves.

## THE PRESENT STUDY

Despite the ubiquity of judgments and predictions that people make about marital satisfaction and marital stability in everyday life, the foregoing review indicates that the question of whether people can make accurate judgments about the marital satisfaction of others has not been well studied and the accuracy of predictions about marital stability has not been studied at all. Moreover, the research that does exist has not examined the characteristics of those who are accurate judges

and those who are not. Research on the accuracy of clinical judgments suggests that common sense notions of who is most accurate (e.g., professional training will improve the accuracy of judgments), when tested, may be proven wrong. And finally, there is research showing the kinds of cues that observers use in making judgments about marital satisfaction, but we do not know whether those same cues are used in making judgments about marital stability, or whether judgment accuracy is related to the use of different kinds of cues.

In the present study, we tested two basic questions related to marital expertise: (a) Can people make accurate judgments of marital satisfaction and marital stability based on observing samples of couples' marital interactions? (b) Are there certain types of people who are particularly accurate at making these judgments? In addressing these questions, we sought out natural groupings where we thought we might find differences in marital expertise: marital researchers, clinical psychology graduate students, marital therapists, pastoral counselors, newlyweds, people in long-term marriages, and recent divorcé(e)s. Ekman and colleagues (Ekman & O'Sullivan, 1991; Ekman, O'Sullivan, & Frank, 1999) have used a similar profession-based procedure in their work on individual differences in the ability to detect lying. Our groups were selected so as to differ in terms of two dimensions related to marital functioning: *professional training* and *personal experience*. These group differences were quantified using an independent panel of raters, and ratings on these dimensions were examined in relation to the accuracy of marital judgments. The use of objective raters to measure characteristics of participants or participant groups has a long tradition in the research literature (e.g., Krefting, Berger, & Wallace, 1978; Maurer & Tarulli, 1997; Sulloway, 1996). Cutting across groups, we measured a number of variables that might also be related to marital expertise. These included demographic characteristics (e.g., sex, age), confidence in ratings, and the use of specific perceived behavioral cues.

The design of this study was based on the assumption that brief segments of marital interaction would be adequate for detecting early warning signs of marital distress and divorce. Carrere and Gottman (1999) found that behavioral coding by trained observers based on the first 3 minutes of marital interaction could be used to predict divorce. Ambady and Rosenthal (1992, 1993) showed that people can make a wide range of accurate

social judgments based on quite *thin* samples of behavior (30 seconds or less). We had participants in our study view a *thicker* sample of behavior: the first 3 minutes of videotaped conversations between married couples who were engaged in 15-minute discussions about an area of conflict in their relationship. This longer sample was consistent with other research on divorce prediction (e.g., Carrere & Gottman, 1999) and also with the practice followed in previous social judgment work of using longer slices of behavior initially to establish that an accuracy effect exists before moving on to increasingly shorter samples. After watching each couple's conversation, participants rated the couple's level of marital satisfaction and predicted whether the couple would still be together in the future. The marital satisfaction of the couples had been assessed at the time the interactions were recorded and the couples had been followed longitudinally. Thus, there were data on how satisfied they were at the time of the recording and also whether they stayed married or divorced. This information provided the criteria for objectively determining the accuracy of participants' judgments of marital satisfaction and marital stability.

METHOD

*Participants*

We recruited a total of 177 participants from nine groups that differ in the extent to which marriage had professional and personal meaning and importance. The groups were the following:

1. Individuals in satisfied long-term marriages (*n* = 24; *M* age = 69.5 years, *SD* = 3.3). Participants were in first marriages and had been married for a minimum of 35 years (*M* duration = 46 years). These couples were randomly selected from couples who had participated in a previous study of marital interaction in long-term marriage conducted in our laboratory (see Levenson, Carstensen, & Gottman, 1993, for a complete description of the three-stage recruitment procedure for the previous study). Individuals in this group had to have marital satisfaction scores (average on two well-established inventories: Burgess, Locke, & Thomes, 1971; Locke & Wallace, 1959) of at least 115.1, which was the satisfaction cutoff score used by Levenson et al. (1993). For the group, the av-

erage marital satisfaction was 123.8 ($SD$ = 5.47).

2. Individuals in dissatisfied long-term marriages ($n$ = 24; $M$ age = 68.8 years, $SD$ = 2.9). Participants were in long-term first marriages (mean duration = 47 years) and were recruited in the same manner as the previous group except that they had to have marital satisfaction scores of less than 115.1. For the group, the average marital satisfaction was 96.7 ($SD$ = 3.76).

3. Newlyweds ($n$ = 16; $M$ age = 28.5 years, $SD$ = 3.4). Participants were in first marriages and had to have been married within the past year. This group was recruited through a local newspaper advertisement that asked for "newlyweds married for the first time within the past 12 months."

4. Recent divorcé(e)s ($n$ = 12; $M$ age = 35.9 years, $SD$ = 8.4). Individuals who had initiated legal divorce procedures or had obtained a legal divorce within the past 2 years were eligible for participation. Participants were recruited through a local divorce attorney and through a newspaper advertisement that asked for "recent divorcé(e)s married only once for 5 years or more, divorced in the last 24 months, and not remarried."

5. Marital therapists ($n$ = 21; $M$ age = 51.2 years, $SD$ = 6.9). Licensed therapists with PhD, MFCC, or MSW degrees and with specialization in marital or couple therapy were recruited through the local chapter of the California Association of Marriage and Family Therapists. Eligible participants had been in clinical practice for a minimum of 5 years. The therapeutic orientation of this group was very diverse, including psychodynamic, systems, cognitive-behavioral, existential-humanistic, gestalt, and eclectic approaches.

6. Marital researchers ($n$ = 13; $M$ age = 45.1 years, $SD$ = 10.1). Researchers with published articles on marriage in scholarly journals of the behavioral or social sciences were contacted and invited to participate. Most participants in this group held academic positions in university departments of psychology.

7. Pastoral counselors ($n$ = 16; $M$ age = 50.7 years, $SD$ = 10.2). This group included Catholic priests, Lutheran reverends, and Jewish rabbis. Eligible participants had practiced marital counseling as part of their work for a minimum of 5 years. Pastoral counselors were recruited using directories of local religious congregations.

8. Graduate students in clinical psychology ($n$ = 20; $M$ age = 29.7 years, $SD$ = 5.9). Graduate students in two PhD programs of clinical psychology (i.e., University of California—Berkeley and the Wright Institute) were recruited. Participants ranged from first- to fifth-year in their programs.

9. Undergraduates ($n$ = 31; $M$ age = 20.7 years, $SD$ = 3.0). To anchor the continuum of personal relevance of marriage, we recruited unmarried college students from the University of California—Berkeley. This group was included so that we had a group with no specific professional training or personal experience related to marital functioning.

## Materials

Participants viewed a videotape that consisted of the first 3 minutes of marital interaction from each of 10 couples. Each couple had been videotaped in our laboratory 6–13 years earlier while having an unrehearsed 15-minute conversation about a mutually identified problem area in their marriage (for a full description of these marital interaction procedures, see Levenson, Carstensen, & Gottman, 1994). These kinds of discussions have proved to be very useful in our previous studies of affective and physiological predictors of current and future levels of marital satisfaction (e.g., Levenson & Gottman, 1983, 1985) and divorce (Gottman & Levenson, 1992, 1999). The recordings were made in a split screen format so that a frontal view of each spouse's head and upper torso was obtained. Individual microphones worn by each spouse enabled a high-quality sound recording.

The selection of the 10 couples for the stimulus videotape was based on three sets of criteria. First, in order to match the current base rate for divorce in the United States (i.e., 50%–67%; see Martin & Bumpass, 1989), we included five couples who, 6 years after the conversation was recorded, were divorced and five couples who were still married. Second, we selected couples whose marital stability was moderately difficult to predict. To determine this, we conducted a pilot study in which we showed videotaped marital interactions of a total of 20 married couples to undergraduate students ($N$ = 116) who indicated whether they thought the couples had stayed together or divorced. Based on the results from this pilot study,

we selected couples for whom the mean accuracy rate of predicting divorce fell between 25% and 75%. In other words, we omitted couples who were either extremely difficult or extremely easy for untrained observers to judge in terms of marital stability. Third, the final set of 10 couples was chosen on the basis of spouses' self-reported marital satisfaction at the time of the interactions (based on the same measures described earlier). To avoid confounding marital satisfaction with marital stability, we included couples with both high and low marital satisfaction in the groups that divorced and that were still married.

### Procedures

Participants were provided with a package consisting of a videotape and a booklet for recording their written responses. The experiment was self-administering and participants were allowed to view the tape at any quiet location that was convenient for them. Participants were told that they would see 10 brief excerpts from conversations of a sample of married couples and were asked to watch each excerpt only once. They were informed that the conversations involved discussion of an important area of disagreement in each marriage and that some of the couples had divorced in the years following the conversation. They were told that the sample of couples was "representative," but the base rate for divorce in the sample was not explicitly stated. After viewing each conversation, participants rated the couple's level of marital satisfaction using a 7-point Likert scale (1 = *very dissatisfied*, 4 = *neutral*, 7 = *very satisfied*) and predicted whether they thought the couple would stay married or get divorced in the future (dichotomous choice). The response booklet was designed so that after participants gave their initial ratings, there was a space to enter revised responses if they changed their mind (this enabled us to preserve their initial choices). Participants also rated their level of confidence in making each of their judgments, again using a 7-point Likert scale (1 = *not at all confident,* 4 = *moderately confident*, 7 = *extremely confident*).

After viewing and rating all 10 couples on the videotape, participants answered several additional questions: (a) their *overall* confidence in judging marital satisfaction and predicting marital stability across all of the couples; (b) an open-ended question concerning the most important qualities (or cues) that influenced the two types of judgments (separately for each type); (c) their estimate

of the percent of all marriages in the United States that end in divorce; and (d) demographic information including sex, age, level of education, marital status, relationship length and satisfaction (for those in a romantic relationship), and marital status of their parents.

Participants were paid $5.00 for their participation and were given the opportunity to receive a report of their overall accuracy in predicting marital stability.

### Panel Ratings

Ratings of the *professional* and *personal* meaning of marriage for our nine groups of participants were obtained from an independent panel of raters, which was composed of 10 doctoral students in psychology. Doctoral students in psychology were selected on the assumption that they would be able to make more informed and differentiated ratings of our nine groups than would other types of raters (e.g., laypersons). However, doctoral students in the subdiscipline of clinical psychology were excluded from the panel in order to avoid systematic bias in the ratings, which might have resulted from having members of one of our groups rate their own group.

Using a 9-point Likert scale (1 = *very little*, 9 = *a great deal*), raters answered the following two questions for each group: (a) "To what extent do you think the understanding of marital relationships would have professional meaning and importance for each of the following groups?" (b) "To what extent do you think the understanding of marital relationships would have personal meaning and importance for each of the following groups?" Interrater reliability estimates on these dimension ratings, based on intraclass correlation coefficients, were 0.96 and 0.92, respectively (Shrout & Fleiss, 1979). As expected, the two dimensions were not significantly correlated with one another ($r = 0.02$). We anticipated that these two dimensions—the professional and personal salience of marriage—would predict group differences in marital expertise. Mean panel ratings for each group on these dimensions are given in Table 1.

### RESULTS

### Who Can Predict Marital Stability?

With 10 couples, accuracy scores could range from 0 to 10 reflecting the number of couples for whom the participant correctly predicted whether

TABLE 1. PANEL RATINGS OF THE PROFESSIONAL AND PERSONAL SALIENCE OF MARRIAGE TO EACH GROUP

| | Professional Salience | | | Personal Salience | | |
|---|---|---|---|---|---|---|
| Rank | Group | *M* | *SD* | Group | *M* | *SD* |
| 1 | Marital researchers | 8.9 | (0.32) | Recent divorcé(e)s | 8.7 | (0.68) |
| 2 | Marital therapists | 8.8 | (0.42) | Dissatisfied long-term marriages | 8.3 | (0.82) |
| 3 | Pastoral counselors | 7.7 | (1.16) | Satisfied long-term marriages | 8.1 | (0.88) |
| 4 | Clinical graduate students | 7.1 | (1.60) | Newlyweds | 8.0 | (1.16) |
| 5 | Recent divorcé(e)s | 5.2 | (2.66) | Marital researchers | 7.5 | (1.25) |
| 6 | Dissatisfied long-term marriages | 4.8 | (2.74) | Marital therapists | 7.0 | (0.68) |
| 7 | Satisfied long-term marriages | 4.1 | (2.60) | Clinical graduate students | 7.0 | (1.25) |
| 8 | Newlyweds | 4.0 | (2.26) | Pastoral counselors | 6.8 | (1.08) |
| 9 | Undergraduates | 2.2 | (2.44) | Undergraduates | 5.5 | (1.65) |

Note: $N = 177$. Ratings of the extent to which marriage holds (a) professional meaning and importance and (b) personal meaning and importance to each group were made on a 9-point scale (1 = *very little*; 9 = *a great deal*).

they stayed together or divorced. The mean overall accuracy for the entire sample of participants was 53.8% correct (range = 20%–80%; $SD = 12.9$). Based on a nonparametric test of proportions, this is significantly different from chance (with chance defined as 50%; $z = 3.16$, $p < .01$). Exploring predictions in terms of the level of marital satisfaction of the target couples, we found that participants made significantly more predictions for divorce for the five couples with the lowest levels of marital satisfaction than for the five couples with the highest levels of marital satisfaction, $t(176) = 6.36$, $p < .001$, $d = 0.96$. However, participants were significantly more accurate in predicting marital stability for the five couples with the highest marital satisfaction than for the five couples with the lowest marital satisfaction, $t(176) = 4.65$, $p < .001$, $d = 0.70$.

*Marital expertise groups.* Table 2 lists mean pre-

diction accuracy (as percent of correct predictions) by group. Based on an omnibus one-way analysis of variance (ANOVA), we found no overall group differences in prediction accuracy. We then tested for group differences in accuracy as a function of the two dimensions of interest: professional and personal salience of marriage (as rated by our independent panel). More specifically, we examined the relationship between the salience ratings of the groups and the accuracy scores of group members by computing two linear trend analyses using ANOVA. In the first analysis, groups were ordered by their mean professional salience rating, and in the second analysis groups were ordered by their mean personal salience rating. There was no significant linear effect for the professional salience or personal salience of marriage on accuracy in predicting marital stability.

We adopted one other method to evaluate group differences in accuracy—directly compar-

TABLE 2. MEAN ACCURACY FOR JUDGMENTS OF MARITAL STABILITY AND MARITAL SATISFACTION

| | Marital Stability | | | Marital Satisfaction | | |
|---|---|---|---|---|---|---|
| Rank | Group | $M^a$ | *SD* | Group | $M^b$ | *SD* |
| 1 | Marital therapists | 57.1 | (10.6) | Satisfied long-term marriages | 0.69 | (0.19) |
| 2 | Undergraduates | 56.8 | (13.3) | Dissatisfied long-term marriages | 0.65 | (0.17) |
| 3 | Recent divorcé(e)s | 56.7 | (10.7) | Newlyweds | 0.63 | (0.17) |
| 4 | Newlyweds | 55.6 | (14.1) | Recent divorcé(e)s | 0.63 | (0.23) |
| 5 | Marital researchers | 55.4 | (12.7) | Pastoral counselors | 0.60 | (0.24) |
| 6 | Clinical graduate students | 52.5 | (14.8) | Marital therapists | 0.59 | (0.13) |
| 7 | Dissatisfied long-term marriages | 51.3 | (13.4) | Clinical graduate students | 0.57 | (0.21) |
| 8 | Pastoral counselors | 50.0 | (14.1) | Undergraduates | 0.57 | (0.23) |
| 9 | Satisfied long-term marriages | 49.2 | (11.0) | Marital researchers | 0.53 | (0.23) |

Note: $N = 177$.
[a]Accuracy scores for predicting marital stability reflect the percent of correct predictions out of the total predictions made (possible range = 0%–100%). [b]Accuracy scores for judging marital satisfaction reflect the overall closeness between participants' standardized ratings of marital satisfaction and the couples' own ratings of their marital satisfaction. For each participant, the mean absolute difference between these ratings was subtracted from a constant (1.41) so that higher scores reflect greater closeness and accuracy (range = 0–1.01).

ing the professional and personal salience of understanding marriage. The set of four groups rated highest in professional salience of marriage (marital researchers, marital therapists, pastoral counselors, and clinical graduate students) was compared to the set of four groups rated highest in personal salience of marriage (recent divorcé[e]s, individuals in dissatisfied and satisfied long-term marriages, and newlyweds). Based on a *t* test, there was no difference in mean accuracy for these two sets of groups in predicting marital stability.

*Demographic characteristics.* Demographic variables were generally not associated with accuracy in predicting marital stability. For sex, comparisons by *t* tests between men and women revealed no differences in accuracy. Correlations between accuracy and the following variables were also nonsignificant: (a) years of education, (b) length of current relationship, (c) participant's current relationship satisfaction, and (d) divorce rate misestimate (i.e., the absolute difference of the participant's estimate from 50%). A regression analysis revealed no significant relationship between prediction accuracy and current relationship status (i.e., single, unmarried committed relationship, married, separated or divorced) with relationship status represented by coded variables.

Two demographic variables were significantly associated with the ability to predict marital stability. First, greater age was associated with lower accuracy ($r = -0.24$, $p = .002$). To help us understand this finding, we conducted two post hoc analyses that revealed that older individuals estimated lower divorce rates in the general population ($r = -0.24$, $p = .001$) and estimated more couples in the sample to be still married ($r = 0.20$, $p = .008$). Second, there was a significant difference in prediction accuracy related to parents' marital status (i.e., married vs. divorced). Specifically, participants whose parents had divorced were more accurate than participants whose parents were still married, $t(169) = -2.29$, $p = .023$, $d = 0.43$. These two groups did not significantly differ in the number of predictions made for divorce, so this cannot account for their differential ability.

*Confidence in predicting marital stability.* Confidence ratings were inconsistently related to accuracy in predicting marital stability. Of the 10 stimulus couples viewed, correlations between confidence and accuracy were significant for eight

TABLE 3. PERCENTAGE OF PARTICIPANTS REPORTING CUE USAGE FOR JUDGMENTS OF MARITAL STABILITY AND MARITAL SATISFACTION

| Cue Category | Marital Stability | Marital Satisfaction |
|---|---|---|
| Communication | 32.2 | 35.6 |
| Positive affect | 29.4 | 37.3 |
| Negative affect | 7.3 | 13.6 |
| Couple traits | 20.9 | 15.3 |
| Nonverbal cues | 12.4 | 29.4 |
| Topic of conversation/other | 17.5 | 5.6 |

Note: $N = 177$.

couples (four correlations were positive and four correlations were negative). Ratings of overall confidence obtained after all couples were rated were unrelated to overall accuracy in predicting marital stability.

*Cues used for predictions.* After rating all of the couples, participants were asked to list the most important qualities that influenced their predictions for marital stability. These open-ended responses were assigned one of 28 codes by two independent coders. Codes included *traits* (e.g., age); *personality*; *similarity*; *"we-ness"* (e.g., partnership, interdependence); *eye contact*; *facial expression*; *body movement*; *conversational tone*; *nonverbal cue* (not otherwise specified); *humor*; *empathy*; *love*; *supportiveness*; *spouses' enjoyment of one another*; *acceptance*; *respect*; *positive affect* (not otherwise specified); *sarcasm*; *anger*; *sadness*; *fear*; *defensiveness*; *disgust/contempt/criticism*; *negative affect* (not otherwise specified); *influence*; *communication*; *topic of conversation*; and *other*. All discrepancies between coders were reconciled such that intercoder agreement became 100%.

Of the 28 perceived cues, only one was significantly correlated with accuracy in predicting marital stability: reported attention to topic of conversation was associated with poorer accuracy ($r = -0.15$, $p = .04$). Cues were also collapsed into six categories: positive affect, negative affect, nonverbal cues, communication, couple traits, and other. A multiple regression analysis with the six cue categories as predictor variables and prediction accuracy as the criterion variable was computed. The final multiple correlation was not significant. Table 3 lists the percentage of participants who reported using these cue categories in their judgments of marital stability and marital satisfaction.

### Who Can Judge Marital Satisfaction?

To determine whether the sample as a whole was more accurate than chance in judging marital satisfaction, we computed a correlation coefficient (*r*) for each participant, reflecting the association between the participant's and couples' ratings of marital satisfaction. Using a Fisher transformation, each *r* was transformed into a *z*. We then computed the mean *z* and transformed this back to the associated *r*. This correlation was significant ($r = 0.48$, $p < .001$), indicating that overall our sample was able to make accurate assessments of marital satisfaction.

For all subsequent analyses involving the accuracy of marital satisfaction ratings, we used a difference score method because it allowed us to examine accuracy in judging individual couples as well as overall accuracy. To derive this difference score measure of accuracy, participants' ratings across the 10 couples were first standardized. Next, the absolute difference between participants' standardized ratings and standardized satisfaction scores for the 10 couples was computed. (It was necessary to standardize because participants' ratings and couples' satisfaction scores were on different scales.) Finally, a mean absolute difference (across the 10 couples) was computed for each participant, and this was subtracted from a constant (i.e., the maximum mean absolute difference in the sample) so that high accuracy would be indicated by high numbers and low accuracy would be indicated by low numbers. The mean accuracy (i.e., the standardized difference score) across all participants was 0.61 (range = 0 to 1.01; $SD = 0.20$).

Due to enduring controversy over the use of difference scores in psychological research (e.g., Rogosa, 1988, 1995), we repeated all analyses involving the overall accuracy of marital satisfaction judgments using accuracy scores based on correlations. Correlation coefficients were computed individually for each participant, reflecting the association between the participant's and couples' ratings of marital satisfaction for the 10 couples. These correlation coefficients (transformed into *z* scores) were used as accuracy scores in this additional set of analyses. Using these correlation-based accuracy scores, the major findings were identical to those we report using the difference scores.

*Marital expertise groups*. Table 2 depicts each group's mean accuracy score for judging marital satisfaction. Based on an omnibus one-way ANOVA, we found no overall differences between groups in accuracy of marital satisfaction judgments. We next examined the linear relationship between judgment accuracy and the professional and personal salience of marriage as rated for each group by our independent panel. Two linear trend analyses were computed using ANOVA, first with groups ordered by their professional salience ratings and second by their personal salience ratings. There was no significant linear effect for the professional salience of marriage on accuracy in judging marital satisfaction. However, there was a significant linear effect for the personal salience of marriage on accuracy in judging marital satisfaction, $F(1,169) = 5.40$, $p = .021$. Thus, the more that understanding marriage had high personal meaning and importance for a group, the more accurate their judgments about marital satisfaction. The four groups rated highest in the personal salience of marriage were recent divorcé(e)s, individuals in dissatisfied marriages, individuals in satisfied marriages, and newlyweds. The same four groups were also most accurate in judging marital satisfaction (see Table 2).

We also directly compared the professional versus personal salience of marriage in terms of judgment accuracy. The set of four groups rated highest in personal salience of marriage were significantly more accurate in judging marital satisfaction than the set of four groups rated highest in professional salience of marriage, $t(144) = 2.59$, $p = .011$, $d = 0.43$.

*Demographic characteristics*. *T* tests revealed no significant differences in accuracy in judging marital satisfaction related to either (a) sex or (b) parents' marital status (i.e., married vs. divorced). Accuracy in judging marital satisfaction was also not related to (a) years of education, (b) current relationship satisfaction, or (c) divorce rate misestimate (i.e., the absolute difference of the participant's estimate from 50%). Based on a regression analysis, current relationship status (i.e., single, unmarried committed relationship, married, separated or divorced) did not explain differences in the ability to judge marital satisfaction. Only two demographic variables were significantly correlated with accuracy: age ($r = 0.18$, $p = .016$) and relationship length ($r = 0.15$, $p = .04$), with increasing values of both associated with greater accuracy in rating marital satisfaction.

*Confidence in judgments*. Reported confidence in

judging seven of the 10 couples was significantly related to accuracy in judging those couples. Confidence and accuracy were positively correlated for ratings of five couples (average $r = 0.21$), but were negatively correlated for ratings of two couples (average $r = -0.23$). In addition, greater overall confidence in judging marital satisfaction across all couples was significantly associated with greater overall accuracy in judging marital satisfaction ($r = 0.25$, $p = .001$).

*Cues used in judgments*. Tested individually, five out of 28 perceived cues were significantly correlated with accuracy in judging marital satisfaction. Specifically, reported attention to personality, sadness, facial expression, and positive affect (not otherwise specified) were all associated with poorer accuracy ($r = -0.18$, $-0.18$, $-0.15$, and $-0.19$, respectively; $p < .05$). Reported attention to disgust/contempt/criticism was associated with greater accuracy ($r = 0.15$, $p < .05$). A multiple regression analysis with six cue categories (positive affect, negative affect, nonverbal cues, communication, couple traits, and other) as predictor variables and rating accuracy as the criterion variable was computed. The final multiple correlation was not significant.

### Judging Satisfaction and Stability: Independent or Not?

To address this question, we examined whether predicting marital stability and judging marital satisfaction are discrete or overlapping abilities. The correlation of the two accuracy scores across the entire sample was positive but not significant ($r = 0.14$, $p = ns$).

### DISCUSSION

### Who Can Predict Marital Stability?

In this study, based on viewing brief segments of marital behavior, participants were able to predict whether couples would stay together or divorce at greater than chance levels. However, the margin of accuracy was low (less than 4% over chance levels). In trying to identify characteristics of the most accurate judges of marital stability, we were most interested in determining whether accuracy was related to the professional and/or personal salience of marriage for our groups. However, neither dimension was found to relate to this type of accuracy. No doubt, this null finding derives in part from lack of variation in accuracy among the nine groups, with all groups hovering within an 8% range around chance levels. Our finding here is quite consistent with much of the literature on clinical judgments, insofar as judgments of trained professionals were no more accurate than untrained laypersons. On the other hand, this result stands in contrast to similar work by Ekman and colleagues who compared members of different occupational groups in terms of their ability to detect deception in 10 targets (Ekman & O'Sullivan, 1991; Ekman et al., 1999). They found that certain groups of professionals (i.e., Secret Service agents and clinical psychologists interested in deception) showed superiority in their task. Thus it appears that the relationship between professional training and accuracy of social judgments may depend both on the type of social judgment and the kind of professional training.

Beyond the professional and personal salience dimensions of our groups, many of the measured characteristics of individual participants also failed to discriminate the most accurate judges of marital stability. Those failing included gender, most other demographic variables, confidence in ratings, and all but one of the perceived cues that participants reported attending to. In fact, only four variables that we tested were related to accuracy in predicting marital stability. First, younger participants were more accurate. Based on follow-up analyses, we believe this may reflect the tendency of our older participants to underestimate the divorce rate both in the sample of couples and currently in the population. Older participants' lower base rates would more accurately reflect those that existed several decades ago. Second, participants whose parents divorced were more accurate. This suggests that the personal experience of family divorce may heighten sensitivity to early indicators of marital break-up. Third, we found that participants who reported attending to the topic of conversation were less accurate. Thus focusing on the content of couples' conflicts when trying to predict the stability of their relationship may be misleading. Fourth, participants were more accurate in predicting stability for couples with higher marital satisfaction than for couples with lower marital satisfaction. We believe this may reflect the important role of external factors in influencing marital stability in distressed couples. Whether happy marriages actually dissolve is highly influenced by such factors as religion, values, children, and economics. In contrast, these kinds of external factors are much less

likely to cause the dissolution of happy marriages. Thus, our observers, who were not privy to these external factors, would not be able to use this information to temper their judgments about the likelihood of dissolution when they viewed couples they considered to be distressed.

Recent research predicting marital stability on the basis of objective coding of emotional behavior has shown a relationship between the emotional content of the first 3 minutes of a marital conversation (as determined by objective coders) and subsequent divorce (Carrere & Gottman, 1999). This finding is strengthened by the present findings insofar as, overall, our participants were able to predict marital stability based on exposure to similar thin slices of behavior at greater than chance levels. However, it does not seem that professional training or personal experience adds much to the accuracy of these predictions.

### Who Can Judge Marital Satisfaction?

Our results indicate that thin slices of behavior are adequate for making accurate judgments as to whether couples are happily married or not. This was indicated by a significant average correlation between participants' ratings of marital satisfaction and the couples' own marital satisfaction scores. However, unlike our findings for judgments of marital stability, there were significant differences in terms of characteristics of our groups of judges. Most importantly, we found that individuals for whom marriage holds high personal meaning and importance, as determined by objective raters, are more accurate in judging how other couples feel about their marriages. This includes individuals who have recently undergone marital transitions—such as recently divorcing or recently marrying for the first time—as well as individuals who have achieved marital milestones such as being in a marriage of greater than three decades duration. Thus it seems that being in a long-lived marriage or going through the intense focus on marital issues that invariably accompanies marital beginnings and endings conveys a sharpened perceptive acuity in judging marital satisfaction in other couples.

Interestingly, the data also suggest that holding some degree of professional training related to marriage does not confer any special advantage in the judgment of marital satisfaction. Groups rated highest in the professional salience of marriage— marital researchers, graduate students in clinical psychology, marital therapists, and pastoral coun-

selors—were also on the low end of accuracy. When this set of professional groups was compared with the set of groups rated highest in personal salience of marriage (recent divorcé[e]s, individuals in dissatisfied and satisfied long-term marriages, and newlyweds), the professional groups were significantly less accurate in assessing marital satisfaction.

These findings raise the obvious question about why certain kinds of professional training related to marriage is associated with lower levels of accuracy in judging marital satisfaction, relative to certain kinds of personal experience. As noted earlier, there have been other studies suggesting that professional training is not associated with superiority in clinical judgment (e.g., see Garb, 1989). Similarly, there have been many studies showing that professional clinical training is not associated with better outcomes in counseling and psychotherapy compared with nonprofessional treatment (for a review, see Christensen & Jacobson, 1994). Our findings that treating or studying marital relationships professionally is associated with *lower* levels of accuracy raises the unsettling possibility that professional training generally does not enhance the ability to assess marital satisfaction accurately.

One more sanguine possible explanation for the low levels of accuracy among marriage professionals in our study is that this results not from training, but rather from the experience of working primarily with distressed couples. Thus, it might be that these professionals have very little experience with satisfied marriages. Interestingly, however, post hoc analyses revealed that the groups with professional training did not estimate lower (or higher) levels of marital satisfaction across couples than did the other groups. Another possibility is that marriage professionals have become accustomed to making judgments based on much more protracted observations of couples' behavior, and thus find it both difficult and unfamiliar to make these judgments based on the kinds of brief samples of marital behavior used in the present study. Support for this interpretation was found in post hoc analyses, which revealed that the four groups with professional training in understanding marriage felt significantly less confident in judging marital stability and marital satisfaction than did the five other groups. Finally, it is important to note that our groups of marriage professionals likely include individuals for whom the personal significance of marriage varies. In future work it will be important to try to disentangle

these two dimensions of salience more completely.

In addition to the group dimensions of professional and personal salience, we also examined a number of characteristics of individual participants. Most of the measured demographic characteristics—including sex, marital status, and satisfaction in one's current romantic relationship—were unrelated to the ability to judge marital satisfaction. None of these were particularly surprising to us, but the lack of sex differences may surprise others. Although women are often thought to be better at decoding interpersonal behavior (e.g., Buck, Miller, & Caul, 1974; Hall, 1984; Hall, Carter, & Horgan, 2000), studies in which the accuracy of social judgments are based on an objective criterion often do not find sex differences. Examples include studies of empathic accuracy in which raters' judgments of what targets are feeling and thinking are evaluated against targets' own reports (Ickes, Stinson, Bissonnette, & Garcia, 1990; Levenson & Ruef, 1992) and studies of deception in which raters' judgments of truthfulness are evaluated against knowledge of whether targets are lying or telling the truth (Ekman & O'Sullivan, 1991). In all of these studies, the accuracy of social judgments made by female and male participants was equivalent.

The individual characteristics that were significantly related to accuracy in judging marital satisfaction all seemed quite expectable. Age and greater length of one's current romantic relationship were both associated with greater accuracy (note that these are somewhat confounded in our study by the inclusion of the two long-term marriage groups, which are limited to first marriages only). Age and marital experience may in fact convey a kind of marital wisdom, perhaps along similar lines to the way that age is associated with increases in crystallized intelligence (i.e., intelligence about practical life matters; e.g., Baltes, Staudinger, & Lindenberger, 1999). Moreover, judges who were more confident in their ratings of marital satisfaction tended to be more accurate in these ratings. This finding is reassuring because it suggests that this is an area where judgments of self-efficacy are related to actual performance, in contrast to our study as well as others of empathic accuracy where confidence in judgments was unrelated to the accuracy of those judgments (Ickes et al., 1990; Levenson & Ruef, 1992). Finally, the use of certain types of perceived cues in judging marital satisfaction was related to accuracy. Specifically, participants who reported attending to

spouses' expressed disgust, contempt, or criticism were more accurate, whereas participants who attended to spouses' personality, sadness, facial expression, or general positive affect were less accurate. This finding is very reminiscent of our previous studies (Gottman & Levenson, 1992; Levenson & Gottman, 1983) where it is the negative emotions and particularly the ''judgmental'' ones (e.g., disgust, contempt, criticism) that are most closely associated with marital distress. Based on these previous studies and the present study of marital expertise, it appears that positive emotions are not as useful in diagnosing the current health of the relationship as are negative ones.

### Marital Expertise: A General Ability?

In this study we examined judgments about two aspects of marital expertise: the ability to predict marital stability and the ability to judge marital satisfaction. In designing this study we did not know if the ability to make these two kinds of judgments would be related. Our findings suggest that these two aspects of marital expertise are orthogonal, a conclusion that is supported by two pieces of evidence. First, there was a nonsignificant correlation between these two abilities. Second, the characteristics of individuals with the ability to predict marital stability were entirely different than those of individuals with the ability to judge marital satisfaction. Thus, it appears that one's level of marital expertise depends in part on the type of marital quality (satisfaction vs. stability) that one is trying to assess. The ability to judge another couple's satisfaction with their relationship does not necessarily imply skill at guessing the likelihood that they will stay together, and vice versa.

### The Difficulty of Predicting Divorce

Although overall our participants were able to predict divorce at better than chance levels, all indications are that this is a very difficult task and that participants (and especially those for whom marriage had personal salience) are better at predicting marital satisfaction than predicting divorce. We were surprised that none of our nine groups predicted marital stability with better than 57.1% accuracy. Perhaps a major reason for the difficulty of predicting divorce is that it is more complex and more multiply determined than marital satisfaction. Although unhappy couples often divorce, many other contextual and cultural fac-

tors (e.g., religion, historical trends, finances, children, ''saving face'') may cause couples to stay together when their marriages clearly are not working. Similarly, some of these contextual factors (e.g., finances, career conflicts) may cause couples who are initially satisfied to eventually divorce. When we view a snapshot of marital interaction (whether in real life or on a videotape), we usually are not aware of contextual factors or of their meaning to the couple. Thus our predictions of stability are based on incomplete data.

### Caveats

There are several important caveats to heed when considering the findings from this research. Although recent research suggests that much shorter samples (i.e., 30 seconds or less) are adequate for making other kinds of social judgments (e.g., Ambady & Rosenthal, 1992, 1993), it may be that judgments of divorce and satisfaction could benefit from being based on longer (and/or multiple) samples of behavior. It is also possible that complex interactions of sample length with professional training could change the patterns of findings for the dimensions of professional and personal salience.

There are also important sampling issues concerning the stimulus materials. It is not clear whether the 10 stimulus couples were representative of the ways that marital satisfaction and marital durability are generally expressed in marital interaction. Also, it was not possible to evaluate whether our participants were indeed representative of the groups they were recruited to characterize. Moreover, it may be that including other groups, such as those with other kinds of specialized training, would provide additional insights. Although we went to considerable lengths to address these issues in our sample selection and participant recruitment, and although most of our findings are consistent with relevant research in the literature (although not always with lay conceptions concerning expertise), it will be important to replicate these findings in independent samples in order to better answer the important question of who the marital experts are.

### Conclusion

This study addressed two primary questions. To the first question of whether observers can predict future marital stability and make accurate judgments about marital satisfaction based on watching brief samples of couples discussing and trying to solve marital problems, our findings suggest that they can. To the second question of whether certain groups of people are more accurate in making these judgments than others, our findings suggest that this depends on which kind of judgment is being assessed. In the case of marital stability, it appears that neither professional training nor personal experience improve the ability to predict divorce, at least for a small slice of behavior. However, for judging marital satisfaction, a class of marital experts is found in those for whom marriage is objectively judged to hold strong personal meaning and importance. Thus, being in a marriage that lasts for three or more decades or experiencing the intense focus on marriage that accompanies divorce or being a newlywed may foster a kind of marital wisdom. These individuals may therefore represent an untapped resource for researchers (e.g., as coders in observational studies of marriage) and for clinicians (e.g., in adding greater understanding about what makes marriages succeed to existing knowledge about what makes marriages fail). This kind of expertise, however, was not found for those who have received some degree of clinical or scientific training related to understanding marital functioning. Finally, across groups, participants who were older, married longer, and who attended to judgmental emotions such as contempt and disgust were most accurate at rating marital satisfaction. On the whole, therefore, it appears that marital expertise in gauging how happy other couples are in their relationships comes with age and marital experience, rather than from professional training.

### NOTE

### REFERENCES

Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin, 111,* 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.

Baltes, P. B., Staudinger, U. M., & Lindenberger, U. (1999). Lifespan psychology: Theory and application

to intellectual functioning. *Annual Review of Psychology, 50,* 471–507.

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.

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

Carrere, S., & Gottman, J. M. (1999). Predicting divorce among newlyweds from the first three minutes of a marital conflict discussion. *Family Process, 38,* 293–301.

Christensen, A., & Jacobson, N. S. (1994). Who (or what) can do psychotherapy: The status and challenge of nonprofessional therapies. *Psychological Science, 5,* 8–14.

Ekman, P., & O'Sullivan, M. (1991). Who can catch a liar? *American Psychologist, 46,* 913–920.

Ekman, P., O'Sullivan, M., & Frank, M. G. (1999). A few can catch a liar. *Psychological Science, 10,* 263–266.

Fincham, F. D., & Bradbury, T. N. (1991). Cognition in marriage: A program of research on attributions. In W. H. Hones & D. Perlman (Eds.), *Advances in personal relationships* (Vol. 2, pp. 159–203). London: Jessica Kingsley.

Floyd, F. J., & Markman, H. J. (1983). Observational bias in spouse observation: Toward a cognitive/behavioral model of marriage. *Journal of Consulting and Clinical Psychology, 51,* 450–457.

Garb, H. N. (1989). Clinical judgment, clinical training, and professional experience. *Psychological Bulletin, 105,* 387–396.

Gottman, J. M. (1998). Psychology and the study of marital processes. *Annual Review of Psychology, 49,* 169–197.

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

Gottman, J. M., & Levenson, R. W. (1999). Rebound from marital conflict and divorce prediction. *Family Processes, 38,* 287–292.

Hall, J. A. (1984). *Nonverbal sex differences: Communication accuracy and expressive style.* Baltimore: Johns Hopkins University Press.

Hall, J. A., Carter, J. D., & Horgan, T. G. (2000). Gender differences in nonverbal communication of emotion. In A. H. Fischer (Ed.), *Gender and emotion: Social psychological perspectives* (pp. 97–117). Cambridge, UK: Cambridge University Press.

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.

Kleinke, C. L., Meeker, F. B., & La Fong, C. (1974). Effects of gaze, touch, and use of name on evaluation of ''engaged'' couples. *Journal of Research in Personality, 7,* 368–373.

Krefting, L. A., Berger, P. K., & Wallace, M. J. (1978). The contribution of sex distribution, job content, and occupational classification to job sextyping: Two studies. *Journal of Vocational Behavior, 13,* 181–191.

Levenson, R. W., Carstensen, L. L., & Gottman, J. M. (1993). Long-term marriage: Age, gender, and satisfaction. *Psychology and Aging, 8,* 301–313.

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

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.

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

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

Margolin, G., Hattem, D., John, R. S., & Yost, K. (1985). Perceptual agreement between spouses and outside observers when coding themselves and a stranger dyad. *Behavioral Assessment, 7,* 235–247.

Martin, T. C., & Bumpass, L. (1989). Recent trends in marital disruption. *Demography, 26,* 37–51.

Maurer, T. J., & Tarulli, B. A. (1997). Managerial work, job analysis, and Holland's RIASEC vocational environment dimensions. *Journal of Vocational Behavior, 50,* 365–381.

Peterson, D. R. (1995). The reflective educator. *American Psychologist, 50,* 975–983.

Raush, H. L., Barry, W. A., Hertel, R. K., & Swain, M. A. (1974). *Communication conflict and marriage.* San Francisco: Jossey-Bass.

Rogosa, D. (1988). Myths about longitudinal research. In K. W. Schaie, R. T. Campbell, W. Meredith, & S. C. Rawlings (Eds.), *Methodological issues in aging research* (pp. 171–209). New York: Springer.

Rogosa, D. (1995). Myths and methods: ''Myths about longitudinal research'' plus supplemental questions. In J. M. Gottman (Ed.), *The analysis of change* (pp. 3–66). Hillsdale, NJ: Lawrence Erlbaum Associates.

Royce, W. S., & Weiss, R. L. (1975). Behavioral cues in the judgment of marital satisfaction: A linear regression analysis. *Journal of Consulting and Clinical Psychology, 43,* 816–824.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86,* 420–428.

Sulloway, F. J. (1996). *Born to rebel: Birth order, family dynamics, and creative lives.* New York: Pantheon Books.

Wills, T. A., Weiss, R. L., & Patterson, G. R. (1974). A behavioral analysis of the determinants of marital satisfaction. *Journal of Consulting and Clinical Psychology, 42,* 802–811.