

## Generalizability of Gottman and Colleagues' Affective Process Models of Couples' Relationship Outcomes

*The generalizability of the affective process models of J. M. Gottman et al. (1998) was examined using a community-based sample of 85 married or cohabiting couples with at-risk backgrounds. Predictive associations between affective processes assessed at about age 21 years and relationship status and satisfaction assessed approximately 2.5 years later were examined. The major findings of Gottman et al. failed to replicate. In particular, men's rejection of their partners' influence, the lack of men's de-escalation of partners' negative affect, and women's negative start-up were not predictive of relationship status. Further, differences in affective processes were found when comparing discussion sessions of the men's versus the women's chosen topics. The findings suggested that the validity and utility of the affective process models need further investigation.*

Affective features of couples' interactions are now recognized as a key component in couples' communication, predicting relationship quality and stability; positive affects are associated with stronger relationship adjustment, whereas negative affects are associated with poorer relationship adjustment or deterioration (e.g., Smith,

Vivian, & O'Leary, 1990; Waldinger, Schulz, Hauser, Allen, & Crowell, 2004). Research on such risk factors has drawn a great deal of attention, particularly from clinicians, because identification of factors that put couples at risk for low levels of satisfaction, separation, or divorce has critical implications for intervention (Heyman & Slep, 2001). The work of Gottman and colleagues has been highly influential in this area (Bradbury & Fincham, 1987; Smith et al., 1990). Unlike many studies that rely on men's and women's self-reports, Gottman's innovative work involved observations of couples' interactions. Further, his work highlighted the importance of dynamic affective processes in such dyads and allowed for close examination of potential mechanisms of relationship maladjustment. The study by Gottman, Coan, Carrère, and Swanson (1998) predicting marital satisfaction and stability (i.e., relationship status) from affect expressed during observed newlywed couples' interactions (referred to in this article as the Newlywed Study) generated much attention, as well as substantial controversy among researchers and practitioners (e.g., DeKay, Greeno, & Houck, 2002; Hafen & Crane, 2003; Heyman & Slep; Stanley, Bradbury, & Markman, 2000).

Gottman et al. (1998) examined a number of affective marital processes, hypothesizing that certain negative affects and sequences of those affects were more destructive than others; they also examined the role of positive affect. They argued that anger may be productive and is less corrosive than affects such as belligerence and defensiveness.

---

Oregon Social Learning Center, 10 Shelton McMurphey Blvd., Eugene, OR 97401 (hyounk@oslsc.org).

*Key Words:* affect, couples interaction, generalizability, process models, relationship satisfaction, relationship stability.

They posited that whereas negative reciprocity is indicative of affective dysregulation, and as such is dysfunctional, escalation of negative affect is a type of power move that indicates rejection of the partner's influence and is particularly deleterious to the future relationship. Following Patterson (1982) and on the basis of their own prior findings, Gottman et al. also hypothesized that the initiation of conflict was a critical issue and that marriages would be more successful if women (who usually were found to initiate conflict in observed laboratory discussions) would soften their start-up by not escalating from neutral to negative during the discussion. Affective patterns expected to predict more positive relationship outcomes included an interaction style whereby one partner de-escalated the conflict by expressing neutral affect following the partner's expression of either high- or low-intensity negative affect. Gottman et al. also posited that positive affect might be predictive of the couple remaining intact and of relationship satisfaction and that the total frequency of positive affect and the ratio of positive to negative affect were indicative of general goodwill toward the partner during the conflict discussion.

Findings in the study of Gottman et al. (1998) partially supported the hypotheses. Anger *alone* was not predictive of relationship outcome. Rather, wife's low-intensity negative affect—defined as anger plus domineering, whining, fear/tension, and sadness—and both husband's and wife's high-intensity negative affect—defined as contempt, belligerence, and defensiveness—predicted greater likelihood of marital separation. Wife's reciprocation of husband's low-intensity negative affect, husband's escalation of wife's negative affect, negative start-up by the wife, lack of husband's de-escalation of wife's low-intensity negative affect, and lower levels of positive affect predicted marital separation. Higher levels of positive affects were predictive of marital satisfaction.

On the basis of these findings, Gottman et al. (1998) argued that marital therapists should tailor their interventions to be more gender specific and should promote interaction styles where wives raise issues more gently and husbands more readily accept influence from wives and de-escalate wives' low-intensity negative affect. To our knowledge, however, no published studies have tested these same or similar models and hypotheses in predicting marital status and satisfaction; thus, the generalizability of these affective process models is in question. It is of particular

concern that careful consideration be given to whether there is strong evidence that recommendations are gender specific or whether similar processes are detrimental to their relationships for both men and women.

A related issue that was raised regarding the gender-specific findings is whether such findings result from the format of the discussion sessions used in the Newlywed Study. Stanley et al. (2000) argued that the pattern of wife's negative start-up and husband's refusal to accept influence from the wife could have resulted from the fact that wives' topics were overrepresented. Gottman, Carrère, Swanson, and Coan (2000) acknowledged that although both the husband and wife were involved in identifying discussion topics, most of the time wives tended to speak first, indicating that it was more likely to be their topic or complaint. Even if women are believed more likely to raise contentious issues in natural conflict situations (Gottman et al., 2000), it is still important to know if dyadic processes differ depending on whether men or women raised the issue (Christensen & Heavey, 1990). In fact, Christensen and Heavey found that the rate of hostile and avoidant behavior from each spouse varied as a function of which partner was requesting change from the other. It is possible that men and women show higher frequencies of negative affect during the discussion of relationship issues that they identified and a higher frequency of positive affect during discussion of the issue identified by their partner.

The purpose of the current study was to examine whether the affective clusters and sequences found to be differentially predictive of relationship status and satisfaction in the Newlywed Study would also be predictive of such outcomes for a community-based sample of 85 married or cohabiting couples with differing characteristics from the sample of Gottman et al. (1998). Affective processes were assessed at about age 21 years, and relationship outcomes were assessed at about age 24 years, approximately 2.5 years later. The study additionally tested the process models while controlling for demographic variables (i.e., marital status—married vs. cohabiting, length of the relationship, presence of children, and socioeconomic status) to account for their possible effects on relationship outcomes. This allowed for examination of whether the effects of the affective process models on relationship outcomes were related to the demographic factors. In addition, the question of whether the prediction patterns varied depending on whose issue was discussed during the

interaction session was examined. This helped to address the issue of whether therapists' recommendations to couples should be gender specific, as advocated by Gottman, or more gender neutral.

With respect to approaches to testing generalizability of findings, in literal replication, the question of whether findings can be duplicated using the same methods of measurement and sampling is tested, whereas a constructive approach to replication goes further in the sense of testing the validity of these methods (Lykken, 1968). That is, use of differing measures and samples has a major advantage in that it demonstrates the robustness and generalizability of the findings (Lykken). The newlywed sample was recruited via news advertising and included newlywed couples with relatively high levels of education (the mean level for husbands and wives was a 4-year college degree). The current study used a sample of young men in an ongoing longitudinal study who were at risk for antisocial behavior (because of residing in higher risk neighborhoods as children) and were recruited by inviting the entire fourth-grade classrooms of boys from randomly selected schools to participate (the Oregon Youth Study). The boys, therefore, were selected to be representative of boys in such neighborhoods in a medium-size metropolitan area. In late adolescence and early adulthood, the young men participated in assessments of couples' interactions with their romantic partners. Given the fact that low-socioeconomic status populations have been underrepresented in research on couples (Karney & Bradbury, 1995), testing the generalizability of the Newlywed Study to couples with lower socioeconomic backgrounds is especially valuable. Relationship status varied from dating to married, as is typical of young adults, and dating couples were eliminated from the current study to test processes for couples who lived together. Although examination of processes only for newlywed couples has the advantage of sample homogeneity, many couples, especially in their 20s (the age of the couples in both the Newlywed and the current studies), cohabit before marriage (Brown, 2000). This is particularly true for those of lower socioeconomic status (White & Rogers, 2000). Therefore, the current study would extend the ecological significance and application of the findings of Gottman et al. if the process models were predictive of relationship outcomes for couples who either cohabit or are married.

The current study involved a similar problem-solving interaction task and the same affect coding used in the Newlywed Study (Specific Affect

Code; Gottman, McCoy, & Coan, 1996). The process models that were tested in the current study included frequency of anger, low- and high-intensity negative affect, negative reciprocity, escalation of negative affect, negative start-up, de-escalation of negative affect, positive affect, and proportion of positive affect to negative affect. Even though men in the current sample were at risk for antisocial behavior when first recruited by virtue of the neighborhoods in which they lived, most did not show elevated behavior problems, and developmental findings for the sample have been comparable to those of other studies in a number of areas (e.g., Capaldi, Crosby, & Stoolmiller, 1996). The reported prevalence of physical aggression for Oregon Youth Study couples at approximately age 18 years was 36% for women and 31% for men, which is comparable to representative national U.S. samples (National Family Violence Survey, Straus & Gelles, 1986, and National Youth Survey, Elliott, Huizinga, & Morse, 1985) and other non-at-risk samples at similar ages (Capaldi & Crosby, 1997). There was also significant variation among the Oregon Youth Study couples in terms of demographic backgrounds (e.g., income and education) and relationship adjustment (e.g., relationship satisfaction and aggression toward a partner). Therefore, it might be expected that similar results to the Newlywed Study would be found with the current sample.

## HYPOTHESES

Hypothesis testing involved prediction from affective interactions assessed at Time 1 (T1, mean age of 21 years) to relationship outcomes at Time 2 (T2, mean age of 24 years). Relationship status was assessed by separation at T2. Levels of relationship satisfaction were examined for couples who remained intact only. For each of the affective clusters and sequences, full definitions are provided in the Method section (Table 1). Detailed below are hypotheses on the basis of findings from the Newlywed Study (a – e), followed by hypotheses unique to the current study.

### *Hypotheses on the Basis of Newlywed Study Findings*

(a) Anger alone at T1 would not be related to relationship status (i.e., separation) or satisfaction at T2. Rather, two composites of negative affect

Table 1. Descriptions of Affect Constructs

Construct	Descriptors
Positive affect (kappa = .80)	
Interest	Elaboration seeking, involved, positive energy
Validation	Understanding, acceptance, paraphrasing
Affection	Tenderness, caring statements, compliments
Humor	Amusement, joking, laughing, wit
Enthusiasm	Happiness, pleasure, anticipation, positive surprise
Low-intensity negative affect (kappa = .54)	
Domineering	Lecturing, patronizing, talking over partner
Whining	Complaining, high-pitched voice
Anger	Irritation, annoyance, raised voice, impatience
Fear/tension	Fidgeting, discomfort, nervous laughter
Sadness	Passivity, sighing, pouting, crying
High-intensity negative affect (kappa = .56)	
Contempt	Sarcasm, insults, hostile humor
Belligerence	Taunting questions, unreciprocated humor, testing limits
Defensiveness	Excuses, countercomplaining, aggressive defense

(low- and high-intensity) at T1 would be predictive of separation at T2 (note that the low-intensity negative affect composite of Gottman et al., 1998, also included anger).

(b) The total frequency of positive affect for men and women and a higher ratio of positive to negative affect at T1 would be predictive of an intact relationship and higher levels of relationship satisfaction at T2.

(c) Reciprocation of negative affect in kind (i.e., one partner's low-intensity negative affect followed by the other partner's low-intensity negative affect and a similar sequence of high-intensity negative affect followed by high-intensity negative affect) at T1 generally would not be significantly related to T2 relationship outcomes, but escalation of partner's negative affect, especially men's escalation of partner's negative affect, would predict separation at T2.

(d) Women's negative start-up, but not men's negative start-up, at T1 would predict separation at T2.

(e) Men's de-escalation, but not women's de-escalation, of partner's negative affect at T1

would predict relationship status (i.e., the couple remaining intact) at T2.

### *Hypotheses Unique to the Current Study*

We hypothesized that men and women would show higher frequencies of negative affect and higher levels of negative start-up during the discussion of the relationship problem issue that they selected. Similarly, we expected that men and women would show a higher frequency of positive affect during discussion of the issue selected by their partner than during discussion of the issue they selected themselves. We also predicted that men and women would make a variety of responses to their partner's criticisms (the latter indexed by frequency of negative affect) and, thus, that men and women would be likely to escalate and to de-escalate negative affect during discussions of the issues their partner selected (rather than just responding with a similar level of negative affect). Given the lack of prior research on the topic of differences in affective process according to whether the problem issue was selected by self or partner, no specific predictions were made regarding which process model would be predictive of relationship outcomes when examined for men's and women's issue sessions separately.

### PLANNED ANALYSIS

Three unique features of the methods employed in the current study (as well as the addition of demographic controls, as previously described) are detailed here:

(a) Whereas the Newlywed Study included only two extreme groups who scored either high or low on the marital satisfaction scale (20 couples for each group) and excluded couples in the midrange, the current study included all married or cohabiting couples regardless of their level of relationship satisfaction. The extreme-group design is known to maximize distinctions between groups, and yet it leads to other concerns such as lack of characterization of the full sample, including those who scored in the midrange on the satisfaction scale, and accompanying limits on generalizability (see Stanley et al., 2000, for further discussions regarding extreme-group analyses). To increase power and to test whether findings would replicate for couples with a range of scores on relationship satisfaction, all cohabiting or married couples in the sample were included in the current study.

(b) Gottman et al. (1998) used joint frequency analyses for the models involving sequential patterns, controlling only for base rates (i.e., overall frequencies) of the consequent behaviors. Such analyses are problematic, however, because in joint frequency analyses, findings are also dependent on the frequency of the antecedent behaviors. Thus, in the current study, frequencies of both antecedent affects and consequent affects were taken into account in the analysis. To illustrate the necessity of controlling for the antecedent affect in addition to the consequent affect, consider cases where one woman reciprocates her partner's negative affect 10% of the total time and another woman reciprocates 20% of the time, which is double the rate and might appear significantly different. If the overall frequency of the antecedent affect—negative affect by the husband—is twice as high in the second couple, however, then the difference in the analyses merely reflects the overall frequency of the men's negative affect.

(c) Some models tested in Gottman et al. (1998) were not tested in the current study, including active listening, contingent positive affect, discriminant functions, and physiological soothing. In the Newlywed Study, active listening was operationally defined as including any sequences of positive affect (interest, affection, humor, validation, and enthusiasm) following a partner's negative affects. As in the Newlywed Study, however, this interaction pattern occurred very rarely during the discussion sessions for our sample. For a discussion of issues regarding the active listening models, see Ginsberg (1999) and Stanley et al. (2000). The contingent positive affect model (a path model) tested by Gottman et al. was not examined in the current study. Although the model presents an interesting mechanism, sequences of behavior in the model were derived from the same coded behaviors; therefore, the value for one variable was highly related to the range of the other variable, and the variables were nonindependent and statistically highly associated. For the same reason, the discriminant function models were not examined. Finally, the physiological soothing model was not tested in the current study because physiological measures were not available.

In the current study, we first conducted logistic regression analyses to examine whether the affective constructs and sequences at T1 predicted relationship outcomes at T2. In predicting relationship status, the couples who separated were compared with couples who stayed together (i.e.,

remained intact). For relationship satisfaction, only couples in an intact relationship were considered, and groups higher or lower in relationship satisfaction at T2 were compared. Therefore, the dependent variables were dichotomous (i.e., intact vs. separated and high vs. low satisfaction), and logistic regressions, which require less stringent assumptions, were used instead of analysis of variance as used in the Newlywed Study. As aforementioned, for models testing affect sequences, frequencies of both the antecedent and the consequent affect were controlled in the logistic regressions. In addition, the same analyses were repeated controlling for marital status (married vs. cohabiting), presence of children in the household, length of the relationship, and socioeconomic status. Finally, the above analyses were repeated examining the couples' interaction sessions led by men and women separately, and we examined whether there were significant differences in affective process depending on whether the problem topic selected by the man or woman was under discussion.

## METHOD

### *Participants*

Male participants were from a community-based sample of 206 young men at risk for delinquency. The Oregon Youth Study is an ongoing multi-agent, multimethod longitudinal study in which participants were recruited through fourth-grade classes (aged 9 – 10 years) in higher crime areas of a medium-size metropolitan region in the Pacific Northwest in 1984 – 1985. The young men have been assessed annually, and the retention rate of the original 206 men was 94% in the 18th year of data collection. The Couples Study is a companion longitudinal study in which the men in the Oregon Youth Study and their partners were assessed three times, namely, in late adolescence for the young men (aged 17 – 20 years), young adulthood (aged 20 – 23 years), and early adulthood (aged 23 – 25 years). For the current study, data from only ages 20 – 23 and 23 – 25 years were used because relatively few couples were married or cohabiting at ages 17 – 20 years, and these older ages are more comparable to those in the Newlywed Study. Note that for clarity of presentation, the two waves used in the current study are designated as T1 and T2. T1 data were collected over 3 years (during Year 12, 13, or 14 of the Oregon Youth Study), depending on when the young man was in a relationship,

and similarly T2 data were collected during Year 15 or 16. A total of 158 couples participated at T1 and 148 couples participated at T2. (Note that there were changes in the young men's relationship status from T1 to T2.) The predominant reason for nonparticipation in the Couples Study at both time points was that they were not in a relationship at the time.

At T1, dating couples ( $N = 72$ , 45.9%) were excluded to increase the comparability to the Newlywed Study by including only married or cohabiting couples ( $N = 85$ , 54.1%), of whom 54 (63.5%) stayed together over the two time points and 31 (36.5%) broke up after the T1 assessment. One additional couple stayed together over time but was excluded from the analyses; observational data were not available for the couple because of poor quality of the video recording. For the couples who stayed together between T1 and T2, the average interval between their T1 and T2 assessments was approximately 2.5 years. The age of men in the current study ranged from 20 to 23.5 years ( $M = 21.3$ ,  $SD = .86$ ) at T1 and from 23 to 25.7 years ( $M = 24.0$ ,  $SD = .55$ ) at T2. The women ranged in age from 16 to 42 years ( $M = 21.2$ ,  $SD = 3.40$ ) at T1 and from 19 to 32.6 years ( $M = 23.7$ ,  $SD = 2.30$ ) at T2.

One man had a 2-year college degree, 48% ( $n = 41$ ) had a high school diploma, 26% ( $n = 22$ ) had a GED or other training experience, and 25% ( $n = 21$ ) had no diploma or GED. Fifty-two percent ( $n = 44$ ) of the women completed high school, 18% ( $n = 15$ ) completed a GED, 4% ( $n = 3$ ) were still in high school, and the other 27% ( $n = 23$ ) had no diploma or GED. At T1, 28 couples (32.9%) were married and 57 (67.1%) were cohabiting. At T2, of those couples still together, 32 couples (59.3%) were married and 22 couples (40.7%) were cohabiting. The average length of the relationship was approximately 2 years at T1 and 4.6 years at T2 for those who stayed with the same partner. At the time of the T1 assessment, 26 couples had a child (or children), and of the 53 couples who stayed together until T2, 39 had at least one child in the home.

### Procedures

Assessments for the Oregon Youth Study included interviews of the parents and young men, questionnaires, interviewer ratings, and court records data. Assessments for the Couples Study included a separate interview and questionnaires for the young men and their partners and

a videotaped session composed of a series of six interactive tasks, including a warm-up discussion, party planning (5 minutes), relationship problem issues (7 minutes for each partner's issue), and goal discussions (5 minutes for each partner's goal). There was a final discussion segment on discipline for couples with children. The couple was seated in chairs close together so that they could touch. The entire couples' assessment lasted approximately 2 hours. To increase comparability to the Newlywed Study, only the two relationship problem discussion sessions were used in the current study. The man and his partner were each directed to pick an issue in their relationship either from the list provided (the Partner Issue Checklist; Capaldi, Wilson, & Collier, 1994) or of their own choosing, and the assessor did not attempt to influence topic choices. Examples of issues selected at T1 included partner's jealousy, not having enough money for dates or activities, and having a hard time talking to each other. The couples discussed each partner's issue for 7 minutes for a total of 14 minutes.

### Measures

*Marital satisfaction.* To assess couples' adjustment and relationship quality, the Dyadic Adjustment Scale (DAS; Spanier, 1976) was used at T2. It consisted of 32 items and has been known to reliably discriminate couples as distressed and non-distressed (Eddy, Heyman, & Weiss, 1991). A cutoff score of 100 on the DAS (Spanier) at T2 was used to form two groups of stable couples who differed in levels of satisfaction. If both partners' scores were 100 or higher, the couple was coded as *stable with high satisfaction* ( $n = 26$ ), and if either partner's score was less than 100, the couple was coded as *stable with low satisfaction* ( $n = 27$ ). The mean couples' satisfaction scores were 118.9 ( $SD = 10.2$ ) for men and 121.1 ( $SD = 9.3$ ) for women in the stable and satisfied couples group and 91.3 ( $SD = 14.5$ ) for men and 94.1 ( $SD = 17.3$ ) for women in the stable but unsatisfied couples group. As described earlier, all married or cohabiting couples were retained in the current study, with the exception of one couple who did not complete the DAS.

*Behavior coding.* The couples' problem discussions were coded with the Specific Affect Code (Gottman et al., 1996), a behavior code designed to detect affect by using indicators of facial expression, vocal tone, physical movement, and

verbal content. A reliable coder previewed the entire videotape of a couple, then watched the tape for a second time, and coded the behavior of the men. Time of onset of each affect was recorded on a computer using the Observational Data Processing Program, software developed at Oregon Social Learning Center (2003). The tape was viewed a third time by the same coder who then recorded the onset of affects by the partner. Six research assistants coded the set of 158 couples tapes. Sixteen percent ( $n = 25$ ) were randomly chosen and assigned as reliability tests. The average kappa (Cohen, 1960) was .74 for both of the passes of coding.

*Affect clusters.* Each coded affect was repeated on a timeline in seconds as in the Newlywed Study. In other words, frequency and duration were identical. With the use of Generalized Sequential Querier (Bakeman & Quera, 1995), the two passes of Specific Affect data (i.e., the man and woman) were merged, and the units of analysis and joint frequencies were computed. As in the Newlywed Study, anger was the only single affect code examined in the analysis. The three clusters or constructs of affects examined were low-intensity negative affect, high-intensity negative affect, and positive affect. Table 1 lists the individual affects included in each construct with descriptors and operational definitions. Kappas were computed again for each affect cluster used in the analyses and are shown in Table 1. The relatively low kappa for the negative affect clusters may be because of the relatively low frequencies of such interaction patterns.

*Affect sequences.* In addition, the following sequences of affects were examined:

(a) Reciprocity of affect: Responding to the partner with an affect with the same affect cluster. Two types of reciprocity were examined, namely, reciprocity of low-intensity negative affect and reciprocity of high-intensity negative affect.

(b) Escalation of negative affect: High-intensity negative affect following partner's low-intensity negative affect.

(c) Negative start-up: Either low- or high-intensity negative affect following partner's neutral affect.

(d) De-escalation: Neutral affect following partner's negative affect (either low or high intensity).

To ensure that each construct and affect sequence was calculated correctly, we followed

the same coding procedures as used in the Newlywed Study and consulted with the second author of that study (Dr. Coan).

## RESULTS

### *Correlations Among the Process Models*

First, correlations between the men's and women's affective behaviors observed during their problem-solving discussions were examined, namely, the frequencies of expressed anger, each of the affect clusters, and the affective sequences of interest (Table 2). Associations between men's and women's scores were generally significant. Expression of positive affect was strongly associated across partners, as was the ratio of positive to negative affect. Frequency of high-intensity negative affect also showed a strong association across partners. Joint frequencies of reciprocation of both high- and low-intensity negative affect were almost perfectly associated across men and women, indicating that these behaviors as coded from continuous parallel streams of affect are almost identical. Men's and women's negative affect reciprocities were examined separately in the remaining analyses, however, because these joint frequencies were analyzed controlling for base rates of antecedent and consequent affects in the logistic regressions and such base rates were different for the man and woman. Because of the likely association of sequential affect scores with the lower frequency affect in the sequence, these associations were examined. The correlations between negative start-up and low-intensity negativity were  $r = .89$  for men

Table 2. *Correlations Between Men's and Women's Scores for Anger, Affective Clusters, and Affective Sequences (N = 85)*

Anger	.43***
Low-intensity negativity	.44***
High-intensity negativity	.68***
Positive affect	.83***
Ratio of positive to negative affect	.83***
Reciprocation of low-intensity negativity	.99***
Reciprocation of high-intensity negativity	.99***
Negative escalation	.41***
Negative start-up	.16
De-escalation of low-intensity negativity	.06
De-escalation of high-intensity negativity	.39***

\*\*\* $p < .001$ .

and  $r = .90$  for women; the correlations between escalation of low-intensity negativity and high-intensity negativity were  $r = .74$  and  $r = .74$  for men and women, respectively; the correlations between de-escalation of low-intensity negativity and low-intensity negativity were  $r = .94$  and  $r = .91$  for men and women, respectively; the correlations between de-escalation of high-intensity negativity and high-intensity negativity were  $r = .92$  and  $r = .84$  for men and women, respectively (all  $p < .001$ ). These strong associa-

tions confirm that affective sequences are highly dependent on the least frequent affect in the sequence; therefore, analyses of sequences should be carefully controlled and interpreted.

*Mean Levels of Affects by T2 Outcome Groupings*

Shown in Table 3 are means and standard deviations for all T1 affect clusters (upper panel) and T1 affect sequences (lower panel) for T2 outcome

Table 3. Mean Levels of Affects and Affect Sequences at T1 by T2 Relationship Outcome Groups

	Intact	Separated	<i>F</i> (1, 83)	High Satisfaction	Low Satisfaction	<i>F</i> (1, 52)
T1 anger, affect clusters, and affect ratios						
Men's anger	9.4 (19.1)	6.7 (9.2)	0.55	5.2 (8.8)	13.8 (25.1)	2.78
Women's anger	14.2 (23.9)	15.9 (29.0)	0.09	10.1 (21.5)	18.2 (26.1)	1.49
Men's low-intensity negativity	118.3 (127.3)	78.6 (61.9)	2.65	75.2 (98.8)	162.9 (139.6)	6.94*
Women's low-intensity negativity	149.6 (121.0)	155.8 (104.9)	0.06	103.4 (100.9)	195.2 (125.5)	8.56**
Men's high-intensity negativity	52.1 (35.9)	48.8 (48.7)	0.13	40.9 (30.1)	62.5 (38.8)	5.09*
Women's high-intensity negativity	49.5 (36.9)	51.9 (47.8)	0.07	37.4 (28.3)	61.7 (41.3)	6.22*
Men's positive affect	72.4 (51.7)	70.4 (58.9)	0.03	97.5 (55.2)	47.6 (35.1)	15.52**
Women's positive affect	73.4 (57.9)	65.8 (51.3)	0.36	100.6 (61.3)	46.1 (41.0)	14.58**
Men's ratio positive/positive + negative affect	0.4 (0.3)	0.4 (0.3)	0.00	0.5 (0.3)	0.2 (0.2)	19.40**
Women's ratio positive/positive + negative affect	0.3 (0.3)	0.3 (0.2)	0.76	0.5 (0.3)	0.2 (0.2)	20.14**
T1 affect sequences						
Men's reciprocity of low-intensity negativity	23.9 (2.7)	24.6 (3.5)	0.03	31.6 (4.7)	22.5 (4.6)	1.81
Women's reciprocity of low-intensity negativity	24.0 (2.6)	25.0 (3.5)	0.05	30.8 (4.6)	23.4 (4.5)	1.22
Men's reciprocity of high-intensity negativity	4.2 (0.5)	5.9 (0.7)	4.19*	4.3 (0.7)	4.2 (0.6)	0.03
Women's reciprocity of high-intensity negativity	3.8 (0.6)	5.6 (0.8)	3.60	3.8 (0.8)	3.9 (0.7)	0.00
Men's negative escalation	10.9 (1.1)	14.5 (1.5)	3.95*	12.9 (1.5)	9.2 (1.4)	2.93
Women's negative escalation	8.7 (0.9)	8.5 (1.2)	0.02	10.9 (1.6)	8.3 (1.5)	1.38
Men's negative start-up	98.1 (4.0)	89.0 (5.3)	1.84	104.1 (6.2)	113.7 (6.1)	1.10
Women's negative start-up	140.6 (4.8)	130.6 (6.4)	1.57	132.3 (6.5)	135.8 (6.4)	0.14
Men's de-escalation of low-intensity negativity	105.1 (3.5)	101.1 (4.6)	0.48	96.6 (5.9)	105.5 (5.8)	1.09
Women's de-escalation of low-intensity negativity	63.3 (2.6)	62.4 (3.5)	0.04	69.4 (4.4)	79.0 (4.3)	2.25
Men's de-escalation of high-intensity negativity	33.4 (1.1)	32.6 (1.5)	0.22	31.5 (1.6)	33.0 (1.6)	0.38
Women's de-escalation of high-intensity negativity	31.8 (1.5)	25.4 (2.0)	6.30*	31.0 (1.7)	34.3 (1.6)	1.93

Note: T1 = Time 1; T2 = Time 2. Means are in seconds. For affect sequences, adjusted means with standard errors were presented.  $n = 54$  for intact group,  $n = 31$  for separated group,  $n = 27$  for low-satisfaction group, and  $n = 26$  for high-satisfaction group. \* $p < .05$ . \*\* $p < .01$ . \*\*\* $p < .001$ .

groups. Group means were compared by one-way analyses of variance. In terms of relationship status at T2 (i.e., intact vs. separated), overall, anger, the affective clusters, and the ratio of positive to negative affect at T1 for men or women were not significantly different across the groups. Of the 12 affect sequences examined, only men’s reciprocity of high-intensity negative affect, men’s escalation of negative affect, and women’s de-escalation of high-intensity negative affect were significantly different between the intact and the separated groups. With respect to future satisfaction for intact couples, all the affect clusters and the ratio of positive to negative affect for both men and women were significantly different between the groups. None of the affective sequences, however, were significantly different between the high- and the low-satisfaction groups.

*Prediction From Anger to Relationship Outcomes*

Following Gottman et al., we predicted that anger alone at T1 would not be related to relationship separation or low satisfaction at T2 but that the two composites of negative affect (low- and high-intensity) at T1 would be predictive of separation but not low satisfaction at T2 (Hypothesis a). Shown in Table 4 are the results of univariate logistic regression analyses predicting separation and relationship satisfaction at T2 from the hypothesized affects at T1. As predicted, neither men’s nor women’s anger alone was significantly predictive of either separation by T2 or relationship satisfaction for couples who remained intact. Contrary to prediction, however, levels of low- and high-intensity negative affect were not significantly predictive of separation

Table 4. *Logistic Regressions Predicting T2 Relationship Outcomes From T1 Affects*

	Relationship Status (T2) (n = 85)		Relationship Satisfaction (T2) (n = 54)	
	B	SE	B	SE
<b>T1 anger, affect clusters, and affect ratios</b>				
Men’s anger	-.012	.016	.034	.024
Women’s anger	.003	.009	.016	.014
Men’s low-intensity negativity	-.004	.003	.007*	.003
Women’s low-intensity negativity	.000	.002	.007*	.003
Men’s high-intensity negativity	-.002	.006	.020*	.010
Women’s high-intensity negativity	.001	.006	.020*	.009
Men’s positive affect	-.001	.004	-.025**	.008
Women’s positive affect	-.003	.004	-.020**	.007
Men’s positive/positive + negative affect	.027	.827	-4.980**	1.490
Women’s positive/positive + negative affect	-.753	.863	-4.730***	1.409
<b>T1 affect sequences</b>				
Men’s reciprocity of low-intensity negativity	.003	.014	-.018	.015
Women’s reciprocity of low-intensity negativity	.004	.015	-.014	.017
Men’s reciprocity of high-intensity negativity	.127*	.065	-.018	.097
Women’s reciprocity of high-intensity negativity	.105	.058	.007	.084
Men’s negative escalation	.063	.033	-.083	.048
Women’s negative escalation	-.004	.040	-.048	.043
Men’s negative start-up	-.014	.009	.010	.012
Women’s negative start-up	-.008	.007	.002	.011
Men’s de-escalation of low-intensity negativity	-.007	.009	.011	.012
Women’s de-escalation of low-intensity negativity	-.003	.015	.022	.015
Men’s de-escalation of high-intensity negativity	-.013	.028	.022	.040
Women’s de-escalation of high-intensity negativity	-.071*	.031	.055	.039

Note: T1 = Time 1; T2 = Time 2. For relationship status prediction, remaining intact was coded as 0 and separated as 1. For relationship satisfaction prediction, high satisfaction was coded as 0 and low satisfaction was coded as 1.

\*p < .05. \*\*p < .01. \*\*\*p < .001.

by T2 but were predictive of low relationship satisfaction.

*Prediction From Positive Affect to  
Relationship Outcomes*

We predicted that the total frequency of positive affect for men and women at T1 and a higher ratio of positive to negative affect would predict the couple remaining intact and higher levels of relationship satisfaction at T2 (Hypothesis b). Findings from the logistic regression analyses (Table 4), however, indicated that neither the total amount of positive affect nor the ratio of positive to negative affect for either men or women was significantly related to the relationship remaining intact. In predicting relationship satisfaction, however, both the total frequency and the ratio of positive to negative affect for men and women were significantly predictive of higher relationship satisfaction.

*Prediction From Sequential  
Affective Processes*

Also shown in Table 4 are the findings from the logistic regressions predicting from affect sequences at T1 to the relationship outcomes at T2 (lower panel).

*Reciprocity and escalation of negative affect.* We predicted that reciprocation of negative affect in kind, specifically one partner's low-intensity negative affect followed by the other partner's low-intensity negative affect and a similar sequence for high-intensity negative affect at T1, would not be significantly related to T2 relationship outcomes but rather that escalation of partner's negative affect from low- to high-intensity, especially men's escalation of partner's negative affect, would predict separation by T2 (Hypothesis c). Findings indicate that contrary to prediction, men's reciprocity of high-intensity negative affect was predictive of separation. Moreover, neither men's nor women's escalation of negative affect at T1 was predictive of separation at T2, and escalation was not predictive of low relationship satisfaction for intact couples.

*Negative start-up.* We predicted that women's negative start-up, but not men's negative start-up, at T1 (i.e., negative affect following partner's neutral affect) would predict separation at T2

(Hypothesis d). Contrary to prediction, findings indicate that neither men's nor women's negative start-up was predictive of separation at T2. Negative start-up also was not related to low relationship satisfaction for intact couples at T2.

*De-escalation of negative affect.* Following Gottman et al., we predicted that men's de-escalation, but not women's de-escalation, of partner's negative affect at T1 would predict an intact relationship at T2 (Hypothesis e). Findings indicate that contrary to prediction, men's de-escalation of their partner's negative affect was not predictive of the relationship remaining intact. Women's de-escalation of their partner's high-intensity negative affect, however, was predictive of remaining intact.

*Demographic controls.* After controlling for demographic variables (i.e., marital status, length of the relationship, presence of child[ren], and socioeconomic status) in the logistic regressions, the findings did not change significantly except for two cases. Women's reciprocity of the partner's high-intensity negative affect was now significantly predictive of separation at T2, and men's escalation of women's negative affect was also predictive of separation.

*Extreme-group analyses.* To examine whether the findings reported above would hold when extreme-group design was used as in the Newlywed Study, the same analyses were repeated with only the 30 couples who scored either high or low on the marital satisfaction scale (15 couples in each group). All the findings from the extreme-group design were the same as in the full sample analyses except for one case: Both men's and women's anger were significantly predictive of low satisfaction.

*Affect in Discussions of Men's Versus  
Women's Issues*

The findings presented above were on the basis of analyses using the two problem-solving discussion sessions combined. That is, the two contiguous discussion sessions (the first addressing the man's chosen issue and the second the woman's issue) were combined and analyzed as a continuous discussion session. We expected that men's and women's affective expressions would vary significantly according to who raised the issue.

To examine this issue, the mean levels of affective expression for men and women were compared for all the affective models previously examined.

As indicated in Table 5, there were significant differences for some affective processes across the two discussion sessions. As hypothesized, women showed a higher frequency of low-intensity (although not high-intensity) negative affect when the problem issues they selected were being discussed. Contrary to hypothesis, however, men did not show higher frequencies of negative affect when the issues they selected were being discussed but in fact showed higher frequency of high-intensity negative affect during discussion of their partners' issues, possibly reflecting defensiveness and escalation. Women, but not men, did show higher frequencies of positive affect during discussion of the problem issues their partners selected than during discussion of the issue they selected. Regarding responses to partner's behaviors (sequential analyses; lower panel), a significant difference was found in

women's reciprocity of low-intensity negative affect. Women showed higher levels of reciprocation of the partner's low-intensity negative affect during the discussion of issues they selected. In addition, significant differences in both negative start-up and de-escalation were found across discussions of the men's versus women's issues. As hypothesized, men showed higher levels of escalation during the discussion of the issues selected by their partners, and women (but not men) showed significantly higher frequencies of negative start-up during the discussion of the issue they selected. Also as hypothesized, men were more likely to de-escalate their partner's low (but not high) -intensity negative affect during discussion of her issue. Contrary to hypothesis, however, women did not show higher frequencies of de-escalation during discussion of issues selected by the men but showed higher frequency of de-escalation of partner's high-intensity negative affect during discussion of the issue they selected themselves.

Table 5. Mean Comparisons of Process Models by Trial (N = 85)

	Men's Issue	Women's Issue	<i>t</i>
T1 anger, affect clusters, and affect ratios			
Men's anger	3.6 (8.6)	4.8 (10.1)	-1.17
Women's anger	6.5 (13.2)	8.3 (15.2)	-1.30
Men's low-intensity negativity	51.0 (55.6)	52.9 (62.8)	-0.39
Women's low-intensity negativity	66.1 (60.2)	85.8 (67.6)	-3.20**
Men's high-intensity negativity	21.2 (21.1)	29.8 (28.2)	-2.76**
Women's high-intensity negativity	26.3 (26.2)	24.1 (20.4)	0.87
Men's positive affect	34.8 (27.7)	36.9 (32.2)	-0.74
Women's positive affect	38.9 (31.8)	31.8 (28.4)	2.77**
Men's positive/positive + negative affect	0.4 (0.3)	0.4 (0.3)	1.13
Women's positive/positive + negative affect	0.4 (0.4)	0.3 (0.3)	3.20**
T1 affect sequences			
Men's reciprocity of low-intensity negativity	10.6 (16.6)	13.6 (23.7)	-1.52
Women's reciprocity of low-intensity negativity	10.3 (15.7)	14.1 (24.1)	-2.01*
Men's reciprocity of high-intensity negativity	2.5 (4.8)	2.3 (3.7)	0.49
Women's reciprocity of high-intensity negativity	2.5 (4.6)	2.2 (3.6)	0.12
Men's negative escalation	4.9 (6.9)	7.3 (10.0)	-2.37*
Women's negative escalation	4.8 (8.9)	3.7 (5.4)	1.58
Men's negative start-up	44.6 (43.3)	50.2 (43.5)	-1.40
Women's negative start-up	62.2 (48.6)	74.8 (54.0)	-2.81**
Men's de-escalation of low-intensity negativity	44.9 (42.9)	58.9 (48.4)	-2.93**
Women's de-escalation of low-intensity negativity	32.0 (37.7)	31.0 (37.9)	0.30
Men's de-escalation of high-intensity negativity	17.0 (15.7)	16.1 (14.5)	0.45
Women's de-escalation of high-intensity negativity	12.0 (12.3)	17.5 (18.9)	-2.59*

Note: T1 = Time 1.

\**p* < .05. \*\**p* < .01.

*Prediction models for men's and women's issues.* Next, we examined each logistic regression prediction model (as examined in Table 4) by T2 outcome groups, namely, intact versus separated and high satisfaction versus low satisfaction, separately for the issues raised by men and women (not shown). All the results for prediction to relationship status were similar to the findings from the combined trials except for a few cases: Men's negative start-up during discussion of their partner's selected issue was now predictive of remaining intact, indicating a counterintuitive finding; women's de-escalation of the partners' high-intensity negative affect remained predictive of remaining intact but only during discussion of the issues they selected; and men's escalation of the partners' negative affect became predictive of separation but only when women's issues were being discussed.

Some differences also emerged in the logistic regression analyses predicting relationship satisfaction at T2 for couples who remained intact. First, men's low- and high-intensity negative affect remained predictive of low satisfaction at T2 only when the issue they selected was discussed, although the latter was only marginally significant. Similarly, women's high-intensity negative affect remained predictive of low satisfaction only during discussion of the problem issue they selected themselves. In addition, when men's issues were being discussed, women's de-escalation of the partner's high-intensity negative affect became predictive of low satisfaction and men's escalation became predictive of high satisfaction, again suggesting somewhat counterintuitive findings. Even after controlling for demographic variables, the findings remained the same except for one case: Men's reciprocity of the partner's high-intensity negative affect during discussion of the issues women selected was now significantly predictive of separation when demographic variables were controlled in the analysis.

## DISCUSSION

The current study aimed to examine the generalizability of the findings of Gottman et al. (1998) for newlyweds regarding the association of affective processes with later relationship outcomes (i.e., relationship status and satisfaction) to young couples from at-risk backgrounds who were either cohabiting or married. In the current study, affective interactions during problem-solving discussions were observed when the young men

and women were about age 21 years and were used to predict to outcomes approximately 2.5 years later. Overall, there was very little predictive association between the affective processes and the relationship status—as indexed by the relationship remaining intact—over the time period studied. Only two of the 22 affective processes examined were significantly associated with remaining intact versus separation: Men's reciprocation of women's high-intensity negative affect was predictive of separation, and women's de-escalation of men's high-intensity negativity was predictive of remaining intact. More association was found between affective processes and later relationship satisfaction for couples who remained intact across the period. All the negative (low- and high-intensity negativity) and positive affect clusters were predictive in the expected direction, as were favorable ratios of positive to negative affect. Neither the individual affect (anger), however, nor any of the affect sequences were predictive of relationship satisfaction. Therefore, with few exceptions, the major findings from the Newlywed Study were not replicated in the current study. A summary of the findings from the two studies (using both the men's and the women's issues combined for the current study) is presented in Table 6.

As discussed, there were some differences between the Newlywed Study and the current study in assessment procedures and in sample characteristics. If the findings were robust and generalizable, however, they would have been expected to replicate in the current study. It thus appears that findings from the study of Gottman et al. (1998) are not robust to differences in sample characteristics or to procedures related to problem-solving discussions. Gottman et al. made very strong suggestions about practice on the basis of their findings, with little consideration to possible limits to applicability to all couples. They argued that therapists should facilitate interaction styles where wives raise issues more gently and husbands more readily accept influence from wives and de-escalate wives' low-intensity negative affect. The findings of the current study, however, call into question the extent to which findings of the Newlywed Study should be used as a basis for recommendations for therapy and interventions with young couples and in general provide a caution against translating empirical findings to treatment recommendations without replications.

One of the major findings of the current study is the failure to provide consistent evidence that

Table 6. Comparisons of the Findings From the Newlywed Study and the Current Study

	Newlywed Study		Current Study	
	Relationship Status	Relationship Satisfaction	Relationship Status	Relationship Satisfaction
T1 affects				
Men's anger	ns	ns	ns	ns
Women's anger	ns	ns	ns	ns
Men's low-intensity negativity	ns	ns	ns	se*
Women's low-intensity negativity	se*	ns	ns	se*
Men's high-intensity negativity	se*	ns	ns	se*
Women's high-intensity negativity	se*	ns	ns	se*
Men's positive affect	se*	se*	ns	se*
Women's positive affect	se*	se*	ns	se*
Men's positive/positive + negative affect	se*	se*	ns	se*
Women's positive/positive + negative affect	se*	se*	ns	se*
T1 affect sequences				
Men's reciprocity of low-intensity negativity	ns	ns	ns	ns
Women's reciprocity of low-intensity negativity	se*	ns	ns	ns
Men's reciprocity of high-intensity negativity	ns	ns	se*	ns
Women's reciprocity of high-intensity negativity	ns	ns	ns	ns
Men's negative escalation	se*	ns	ns	ns
Women's negative escalation	ns	ns	ns	ns
Men's negative start-up	ns	ns	ns	ns
Women's negative start-up	se*	ns	ns	ns
Men's de-escalation of low-intensity negativity	se*	ns	ns	ns
Women's de-escalation of low-intensity negativity	ns	ns	ns	ns
Men's de-escalation of high-intensity negativity	ns	ns	ns	ns
Women's de-escalation of high-intensity negativity	ns	ns	se*	ns

Note: T1 = Time 1. ns represents nonsignificance, and se\* represents significance.

women's negative start-up was a strong predictor of separation. Gottman et al. (1998) argued that a gender-patterned women-demand and men-withdraw interaction style can characterize most of the conflictual situations in couples and that this interaction pattern is detrimental to relationship outcomes. It is possible, however, that such interaction patterns are related to the age and developmental stages of the couple. The dyadic process of young couples may be more volatile, and their power structure within the dyad may not be settled or may be more equal than that for older couples with more established relationships, and thus such patterns may have less predictive power for relationship adjustment. Unexpectedly, men's negative start-up during discussion of the woman's issue predicted the couple remaining *intact*. As this was not hypothesized, and as the discussion time for each individual issue was relatively short, the finding should be interpreted with caution. It

may be, however, that stronger affect on the part of the men at this stage of the relationship in response to their partners' complaints suggests more engagement with and commitment to the partner.

Our findings regarding the role of men's escalation of negative affect (interpreted by Gottman et al., 1998, as men's refusal to accept women's influence) in couples' relationships indicated somewhat inconsistent results. Men's escalation of women's negative affect significantly predicted separation only when women's issues were being discussed, which is consistent with the finding of Gottman et al.; the same interaction pattern was not significant when men's issues were being discussed or when both men's and women's issue sessions were examined together. Gottman and colleagues have long argued that there are gender-asymmetric patterns in couples' interactions (Coan, Gottman, Babcock, & Jacobson, 1997) and that these pertain to power. On the

basis of the work on violent couples, Gottman et al. argued that men would want to maintain their power in the relationship and thus be less likely to accept influence from women and that the latter is a significant predictor for relationship stability. One might argue that our finding is viewed as evidence to confirm the finding of Gottman et al., given that in the study of Gottman et al., women were more likely to initiate the discussion (similar to our women's issue session). As indicated in Table 5, however, women tended to show relatively high levels of low-intensity negative affect when their own issues were being discussed. This relatively high volume of negative affect could be partially responsible for evoking more defensive behavior (i.e., high-intensity negative affect).

Another major finding in the current study is that, overall, both men and women tended to show more negative affect and less positive affect when they discussed issues that the women raised, and there were some differences in prediction patterns when comparing sessions led by men and women. In particular, some of the affect sequences were predictive of relationship status only when issues women selected were being discussed. This indicates that problem-solving interactions do vary according to who raised the problem issue, that procedures involved in generating the discussion topics and the format of the discussion session may be related to couples' interaction patterns, and that this should be considered in studies testing hypotheses regarding couples' interactions. Examination of such differing patterns may be very helpful in understanding negative and positive affective interactions in couples and the association to relationship outcomes and may lead to some gender-specific prevention recommendations. Further investigation of this issue may also illuminate the nature of women-demand and men-withdraw interaction patterns.

Some of the counterintuitive findings suggest that generalizability of meanings (or interpretations) of the process models requires further consideration. For example, during discussion of the men's issues, women in the low-satisfaction group were more likely than women in the higher relationship satisfaction group to de-escalate from partner's high-intensity negative affect. Neutral affect following a partner's negative affect could be a sign that the individual is disengaging for negative reasons rather than a positive interpretation that they are defusing the negative affect. It is possible that some individuals with histories of con-

flict discussions that do not achieve positive resolution find displays of negative affect by their partner aversive or find such discussions very difficult, perhaps because of lower levels of education, and are therefore more likely to give up the attempt to resolve the issues in the relationship, resulting in their "backing off" to neutral affect. Couples' affective processes are likely to have different meanings depending on socioeconomic status, developmental stage, or cultural backgrounds. Therefore, conceptualization of couples' sequential interaction patterns may be quite complex, and a similar pattern may be related to positive patterns in some instances and to negative patterns in others.

The findings of the current study indicate that the role of anger in couple relationships needs further investigation. When the extreme-group design was used for additional tests (similar to the study of Gottman et al.), both men and women in the low-satisfaction group showed significantly higher levels of anger than those in the high-satisfaction group. Given that studies have shown inconsistent findings regarding the effects of anger on relationship quality (see Waldinger et al., 2004), this is not a surprising finding. In fact, the argument of Gottman and Krokoff (1989) that, in the long run, anger was related to improvement in marital satisfaction has also been criticized by many as an artifact (see Jacobson, 1990; Woody & Costanzo, 1990, for further discussions). It is also necessary to consider the role of anger in the context of the specific coding system used in both studies. According to the Specific Affect Coding System, other high-intensity negative affects (e.g., contempt, belligerence) take precedence over anger; thus, the anger code is relatively less intense (Waldinger et al.). Consequently, it is possible that differences between anger and the other negative affects are in fact quantitative rather than qualitative (Waldinger et al.), suggesting that the comparison between anger and the low- and high-intensity negative affect models is not a valid way to examine the relative importance of different negative affects. This issue is exacerbated when anger alone is compared to composite scores of negative affects, because any individual affect code (e.g., anger) has a lower frequency of occurrence than a cluster of codes (e.g., low- and high-intensity negativity). Models examining anger alone are bound to have less power to detect associations with outcomes compared with models involving clusters (Stanley et al., 2000). We believe that the

possibility of differential roles of individual negative affects in the deterioration of couples' relationships is an important issue, but the role of anger versus other affects needs further consideration and empirical testing.

Measurement issues in the operationalization of the process models deserve further discussion. Some of the issues are closely related to the uniqueness of the Specific Affect Coding System. The Specific Affect Code is a two-stream coding system. That is, the affects of both partners are coded on a second-by-second basis, and this is different from other coding systems that involved a single-code stream where only behavior of the speaker is coded. Therefore, with the Specific Affect Coding System, it is more difficult to distinguish action and reaction. This becomes especially problematic when gender differences were tested in sequences that were essentially mutual sequences of affect. Gottman et al. (1998) used sequences lagged by 1 second; therefore, each consequent behavior can also be an antecedent behavior. Thus, when men's reciprocity of partner's negative affect was examined, women's negative affect is the antecedent behavior, but the same women's negative affect can be a consequent behavior when examining women's reciprocity of partner's negative affect. It is particularly problematic when base rates of both antecedent and consequent behaviors were not controlled for as in the Newlywed Study.

The measurement issues suggest that conclusions regarding gender differences from the Newlywed Study should be reexamined. The lack of independence of the observed behaviors of two individuals involved in a dyadic discussion is a challenging issue and requires conceptual and statistical approaches that can accommodate such interactive behavior. Affect is frequently communicated by facial expressions and body language with no verbal expression, and an affective signal may be fleeting. Thus, the Specific Affect Coding System has many advantages and may be more ecologically valid than single-stream codes. The approach does require careful consideration regarding operationalization and interpretation of the affects, however, in particular, the interplay of affect in the dyad. Of course, it is also the case that single-stream codes are assessing an interaction, and many of the same considerations regarding nonindependence of the behaviors of members of the dyad must be considered.

Taken together, we believe that extreme caution should be used in making gender-specific

recommendations to couples regarding their interactions. This concern is on the basis of a number of issues, such as the possible poor replicability of findings regarding gender differences. Further, to test gender differences adequately, the significance of direct comparisons across gender should be tested. Other considerations are that gender issues in behavior are fraught with stereotypical interpretations, which may be counterproductive in achieving well-functioning couple relationships. Even when a significant gender difference in behavior is found, this usually means that there is some tendency to a difference, but for many men and women, there may be no difference, or for some couples the behavior differences may actually be in the opposite direction. Thus, it is questionable how much study findings apply to any individual couple. It may be more productive and accurate to make a recommendation that is gender neutral (e.g., that negative start-up is a behavior that can be destructive in a relationship and that both men and women should try to raise their problem issues in a constructive and nonblaming fashion).

Some of the reasons that the current study failed to replicate major findings of the Newlywed Study may be attributed to differences in procedures, such as design of the discussion tasks (e.g., the role of the assessors in choosing discussion topics in the Newlywed Study, the presence of warm-up sessions before the problem-solving sessions in the current study, and a single session discussing both partners' issues in the Newlywed Study vs. two independent discussion sessions talking about each partner's issue separately in the current study), and physical context of the interaction (wearing physiological sensors in the Newlywed Study vs. a more naturalistic interaction in the current study). These differences could have led to differing affective interaction patterns among the couples. The Newlywed Study also used extreme-group analyses, although reanalyses using the same design in the current study resulted in very similar findings to our original findings, except for the anger model (i.e., both men and women in the low-satisfaction group showed significantly higher levels of anger than those in the high-satisfaction group).

In addition, sample differences could be partially responsible for the divergent findings. Compared to couples in the Newlywed Study, couples in the current study were slightly younger and had significantly lower incomes and educational backgrounds. Also, couples who were

cohabiting and who had children were included in the current study. Cohabiting status, the presence of children, and socioeconomic status are proximal and salient contextual factors for couples (Amato, Johnson, Booth, & Rogers, 2003; Stanley, Whitton, & Markman, 2004); therefore, these factors might be systematically related to interaction patterns among couples, although our further analyses controlling for these factors resulted in very similar findings.

As we focused on generalizability of the findings of Gottman et al. (1998), many of the conceptual and methodological issues in the Newlywed Study were not fully addressed in the current study. First, examining validity of the process models is of paramount importance. The process models of Gottman et al. have a strong intuitive appeal, but the way the process models were operationalized and analyzed was not adequately aligned with the way the observational data were constructed within the Specific Affect Coding System. As described earlier, some of the process models were not conceptually independent and were highly associated statistically. Thus, they could not be considered as competing models. Such an approach obscures the actual meaning of the interactional patterns within dyads.

Second, as mentioned earlier, in sequential behaviors, it is inevitable that there would be more opportunities to drive particular sequences when there are more antecedent behaviors; therefore, antecedent behaviors should be controlled in the analyses. It is very likely that findings from the Newlywed Study were affected by the choice of testing the models, that is, using joint frequencies and controlling only for the frequency of the consequent behavior. Wampold (1989) suggested using nonparametric sequential statistics (e.g., kappa) to study sequential interaction data that are not affected by the frequency of the antecedent and consequent behaviors. In addition, in a recent study, Dagne, Howe, Brown, and Muthén (2002) suggested a new way to analyze sequential data using log-odds ratios and empirical Bayes estimation. Further, survival analysis that takes into account both the frequency of transitions and the duration in state would be a better way to analyze sequential behaviors coded in dual streams, such as Specific Affect data (M. Stoolmiller, personal communication, February 13, 2004; see also Snyder, Stoolmiller, Wilson, & Yamamoto, 2003). Our future work will consider these new methods in testing process models.

The current study is one more step toward the difficult but crucial goal of increased understanding of couples' affective interactions and their association with relationship outcomes. Although the current study features a unique population (at-risk background, White, and lower socioeconomic status) and is expected to contribute to the field, the generalizability of findings from our study also should be tested. Future research should examine process models with samples of diverse backgrounds. Furthermore, in addition to the conceptual and methodological issues discussed above, several other issues remained unanswered. For instance, it is puzzling that some affect models were significantly related to relationship stability but were not related to relationship satisfaction and vice versa. It appears that for relationship satisfaction, the amount of positivity and negativity matters rather than specific interaction patterns within dyads. This could be an indication of different mechanisms that lead to or are indicative of future relationship stability versus relationship satisfaction in young couples. The amount of positive and enjoyable interaction is likely to protect and even increase satisfaction levels over time and the amount of negativity to erode satisfaction. Specific negative interaction patterns, however, could be more indicative of some fundamental differences and disagreements between the couple (e.g., one partner being unable to accept the other's high level of substance use). Few studies have examined differential mechanisms for key adjustment outcomes in couples, and this issue deserves further investigation. In addition, changes in couples' interaction patterns over time and the associations of such change to couples' adjustment are important questions to pursue.

In all, the current study failed to replicate major findings of the study of Gottman et al. (1998) of affective interaction and relationship outcome, similar to two recent studies that have failed to replicate the batterer typology work of Gottman et al. (1995) (Babcock, Green, Webb, & Graham, 2004; Meehan, Holtzworth-Munroe, & Herron, 2001). One might argue that there were several consistent findings between the current study and the Newlywed Study, such as overall nonsignificant group differences in reciprocation of negativity in kind. Such concordance between the two studies in failure to reject null hypotheses, however, is not the same as replicated significant findings. Further efforts to test the role of affect during couple's conflict in predicting relationship

satisfaction and stability are needed before any specific recommendations are made for clinicians. We cannot support the suggestion of Gottman et al. to disregard negative affect reciprocity and to put more focus on men's rejection of their partners' influence (as defined) and women's negative start-up, or any other gender-related implications on the basis of their findings. Some of the process models suggested by Gottman et al. may prove to be very useful tools in clinical settings, but the current study generated many questions about conceptual aspects of those models and their predictive power for relationship outcomes. Validity and utility of the process models await further replication work.

#### NOTES

The Cognitive, Social and Affective Development, National Institute of Child Health and Human Development (NICHD), and Division of Epidemiology, Services and Prevention Branch, National Institute on Drug Abuse (NIDA), National Institutes of Health (NIH), U.S. PHS (Public Health Service) provided support for the Couples Study (Grant HD 46364). Additional support was provided by Grant MH 37940 from the Antisocial and Other Personality Disorders Program, National Institute of Mental Health (NIMH), U.S. PHS; Grant DA 051485 from the Division of Epidemiology, Services and Prevention Branch, NIDA, and Cognitive, Social, and Affective Development, NICHD, NIH, U.S. PHS; and Grant MH 46690 from the Prevention, Early Intervention, and Epidemiology Branch, NIMH, and Office of Research on Minority Health, U.S. PHS.

#### REFERENCES

- Amato, P. R., Johnson, D. R., Booth, A., & Rogers, S. J. (2003). Continuity and change in marital quality between 1980 and 2000. *Journal of Marriage and Family, 65*, 1 – 22.
- Babcock, J. C., Green, C. E., Webb, S. A., & Graham, K. H. (2004). A second failure to replicate the Gottman et al. (1995) typology of men who abuse intimate partners ... and possible reasons why. *Journal of Family Psychology, 18*, 396 – 400.
- Bakeman, R., & Quera, V. (1995). *Analyzing interaction: Sequential analysis with SDIS and GSEQ*. New York: Cambridge University Press.
- Bradbury, T., & Fincham, F. D. (1987). Assessment of affect in marriage. In D. K. O'Leary (Ed.), *Assessment of marital discord: An integration for research and clinical practice* (pp. 59 – 108). Hillsdale, NJ: Erlbaum.
- Brown, S. L. (2000). Union transitions among cohabitators: The role of relationship assessments and expectations. *Journal of Marriage and the Family, 62*, 833 – 846.
- Capaldi, D. M., & Crosby, L. (1997). Observed and reported psychological and physical aggression in young, at-risk couples. *Social Development, 6*, 184 – 206.
- Capaldi, D. M., Crosby, L., & Stoolmiller, M. (1996). Predicting the timing of first sexual intercourse for adolescent males. *Child Development, 67*, 344 – 359.
- Capaldi, D. M., Wilson, J. E., & Collier, M. (1994). *Partner Issues Checklist*. (Unpublished OSLC instrument, available from OSLC, 10 Shelton McMurphy Blvd., Eugene, OR 97401)
- Christensen, A., & Heavey, C. L. (1990). Gender and social structure in the demand/withdraw pattern of marital conflict. *Journal of Personality and Social Psychology, 59*, 73 – 82.
- Coan, J., Gottman, J. M., Babcock, J. C., & Jacobson, N. S. (1997). Battering and the male rejection of influence from women. *Aggressive Behavior, 23*, 375 – 388.
- Cohen, J. A. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*, 37 – 46.
- Dagne, G. A., Howe, G. W., Brown, H. C., & Muthén, B. O. (2002). Hierarchical modeling of sequential behavioral data: An empirical Bayesian approach. *Psychological Methods, 7*, 262 – 280.
- DeKay, M. L., Greeno, C. G., & Houck, P. R. (2002). Searching for a two-factor model of marriage duration: Commentary on Gottman and Levenson. *Family Process, 41*, 97 – 103.
- Eddy, J. M., Heyman, R. E., & Weiss, R. L. (1991). An empirical evaluation of the Dyadic Adjustment Scale: Exploring the differences between marital "satisfaction" and "adjustment." *Behavioral Assessment, 13*, 199 – 220.
- Elliott, D. S., Huizinga, D., & Morse, B. J. (1985). *The dynamics of delinquent behavior: A National Survey Progress Report*. Boulder: Institute of Behavioral Sciences, University of Colorado.
- Ginsberg, B. G. (1999, January 28). *Ginsberg response to Gottman JMF, Feb 1998 article*. Message posted to smartmarriages mail, archived at <http://archives.his.com/smartmarriages/msg00614.html>
- Gottman, J. M., Carrère, S., Swanson, C., & Coan, J. (2000). Reply to "From basic research to interventions." *Journal of Marriage and the Family, 62*, 265 – 273.
- Gottman, J. M., Coan, J., Carrère, S., & Swanson, C. (1998). Predicting marital happiness and stability from newlywed interactions. *Journal of Marriage and the Family, 60*, 5 – 22.

- Gottman, J. M., Jacobson, N. S., Rushe, R. H., Wu Shortt, J., Babcock, J., La Taillade, J. J., et al. (1995). The relationship between heart rate reactivity, emotionally aggressive behavior, and general violence in batterers. *Journal of Family Psychology, 9*, 227 – 248.
- Gottman, J. M., & Krokoff, J. L. (1989). Marital interaction and satisfaction: A longitudinal view. *Journal of Consulting and Clinical Psychology, 57*, 47 – 52.
- Gottman, J. M., McCoy, K., & Coan, J. (1996). The Specific Affect Coding System. In J. M. Gottman (Ed.), *What predicts divorce? The measures* (pp. 1 – 169). Hillsdale, NJ: Erlbaum.
- Hafen, M., Jr., & Crane, D. R. (2003). When marital interaction and intervention researchers arrive at different points of view: The active listening controversy. *Journal of Family Therapy, 25*, 4 – 14.
- Heyman, R. E., & Slep, A. M. S. (2001). The hazards of predicting divorce without crossvalidation. *Journal of Marriage and Family, 63*, 473 – 479.
- Jacobson, N. S. (1990). Contributions from psychology to an understanding of marriage. In F. D. Fincham & T. N. Bradbury (Eds.), *The psychology of marriage* (pp. 258 – 275). New York: Guilford.
- Karney, B. R., & Bradbury, T. N. (1995). The longitudinal course of marital quality and stability: A review of theory, method, and research. *Psychological Bulletin, 118*, 3 – 34.
- Lykken, D. T. (1968). Statistical significance in psychological research. *Psychological Bulletin, 70*, 151 – 159.
- Meehan, J. C., Holtzworth-Munroe, A., & Herron, K. (2001). Maritally violent men's heart rate reactivity to marital interactions: A failure to replicate the Gottman et al. (1995) typology. *Journal of Family Psychology, 15*, 394 – 408.
- Oregon Social Learning Center. (2003). *Observational data processing program*. Unpublished manual. Revised August 2003.
- Patterson, G. R. (1982). *Coercive family process*. Eugene, OR: Castalia.
- Smith, D. A., 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.
- Snyder, J., Stoolmiller, M., Wilson, M., & Yamamoto, M. (2003). Child anger regulation, parental responses to children's anger displays, and early child antisocial behavior. *Social Development, 12*, 335 – 360.
- Spanier, G. B. (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. M., Bradbury, T. N., & Markman, H. J. (2000). Structural flaws in the bridge from basic research on marriage to interventions for couples. *Journal of Marriage and the Family, 62*, 256 – 264.
- Stanley, S. M., Whitton, S. W., & Markman, H. J. (2004). Maybe I do: Interpersonal commitment and premarital or nonmarital cohabitation. *Journal of Family Issues, 25*, 469 – 519.
- Straus, M. A., & Gelles, R. J. (1986). Societal change and change in family violence from 1975 to 1985 as revealed by two national surveys. *Journal of Marriage and the Family, 48*, 465 – 478.
- 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.
- Wampold, B. E. (1989). Kappa as a measure of pattern in sequential data. *Quality and Quantity, 23*, 171 – 187.
- White, L., & Rogers, S. J. (2000). The economic context of relationship outcomes: A decade in review. *Journal of Marriage and the Family, 62*, 1035 – 1051.
- Woody, E. Z., & Costanzo, P. R. (1990). Does marital agony precede marital ecstasy? A comment on Gottman and Krokoff's "marital interaction and satisfaction": A longitudinal view. *Journal of Consulting and Clinical Psychology, 58*, 499 – 501.