Representativeness of Observed Couple Interaction: Couples Can Tell, and It Does Make a Difference

David A. Foster, Robert D. Caplan, and George W. Howe
George Washington University

Research evidence suggests that people's public self-presentations may be biased in socially desirable directions. Using videotaped samples of couple interactions, this study examined the extent to which self-presentational bias occurs in such samples and the impact of such bias on the predictive validity of observed behavior. Each member of 239 couples rated the typicality of their partner's taped socially supportive and undermining behaviors. Separate multi-item, internally consistent measures of typicality of support and undermining were found. Analyses showed that these measures did not appear subject to bias and could identify subgroups that varied notably in the criterion validity of their observational data. These effects appeared for typicality of social support but not for typicality of social undermining. Partner ratings may be a promising tool for isolating invalid samples of observed behavior.

One of the most challenging tasks in psychological research is the collection of valid observational data on people in normal social interactions. Such observations are usually interpreted as reflections of more general patterns of behavior because they occur in natural settings, and they are considered valid if they are such reflections. Scientists who conduct observational research, however, have long been aware that various characteristics of the research design can affect participants' behavior during such data collection. For example, Gottman (1979, p. 248) has found evidence that couples' behavior is more negative, with more negative affect reciprocity and less agreement, when assessed in the home rather than in a laboratory setting.

A large body of evidence suggests that such effects may occur because people tend to bias their self-presentations of private behavior in socially desirable directions during public interactions (e.g., Adler, 1930; Allport, 1955; Crowne & Marlowe, 1964; Leary & Kowalski, 1990; Orne, 1969; Rosenberg, 1979). Almost no research has been conducted, however, to determine whether self-presentation bias does occur in observational samples of behavior and to examine the impact of such bias on the validity of observed measures. The purpose of this study was to address these questions.

To do so, we asked members of couples being observed in an interaction with each other to rate whether their partner's behavior was typical. When we first considered asking partners to report on the representativeness of one another's behavior, we assumed that we would find studies in which this straightforward approach had been used. We found, however, only one such published study that had been conducted within the last

30 years (Warren & Gilner, 1978), and it cited no other similar research. That study of couples found no evidence of atypical partner behavior.

The current study differs from Warren and Gilner's (1978) in three ways. Warren and Gilner had 41 couples listen to and rate how "true to life" their partner's responses were regarding how they would handle 15 hypothetical situations designed to elicit positive assertion in intimate relationships. In the current study, we had each partner rate how representative the other partner's behavior was on the basis of an actual interaction the couple had just completed. Second, Warren and Gilner used a single item to assess representativeness. In this study, we assessed perceptions of typicality using multi-item, internally reliable measures. Third, Warren and Gilner's single-item rating of how "true to life" their partner's responses were precluded ratings of specific dimensions of behavior. In this study, participants rated the representativeness of their partner's behavior along specific dimensions of behavior that have proven important conceptually and empirically in studying couples' adjustment, conflict, and marital stability (e.g., Gottman, 1979; Markman, Renick, Floyd, & Stanley, 1993). These dimensions are Emotional Validation and Emotional Invalidation.

Emotionally validating behavior is defined as actions that show emotional support or concern for the other person, whereas emotionally invalidating behavior involves statements that undermine the other person through criticism, sarcasm, or guilt induction. The concept of emotional validation bears strong similarity to a subset of actions described under the rubric of social support. *Social support* has been defined as subsuming both emotionally validating acts and acts aimed at instrumental aid in coping. In a like fashion, emotionally invalidating behavior is similar to the concept of social undermining.

There is a substantial literature on the effects of recipients' self-reports of received support and undermining on well-being (e.g., Abbey, Abramis, & Caplan, 1985; Cohen & Wills, 1985; Coyne & Bolger, 1990; Coyne & DeLongis, 1986; Rook, 1984; Sarason, Sarason, & Pierce, 1990; Vinokur, Schul, & Caplan,

1987). Literature going back more than 20 years generally supports the conclusion that social support has a beneficial effect on well-being, including the ability to buffer the effects of other stressors on mental and physical health (e.g., Cobb, 1976; Cohen & Wills, 1985; Gerin, Milner, Chawla, & Pickering, 1995; Rhodes & Woods, 1995). Assessments involving actual behavior, however, are the exception rather than the rule (e.g., Karney & Bradbury, 1995; Kashy & Snyder, 1995). Although the recipient's perception of social support is typically the most immediate antecedent of well-being (e.g., Wethington & Kessler, 1986), valid behavioral assessments of support are necessary to understand the link between behavior and perception of support. Further progress in behavioral assessment may depend, in part, on understanding whether observed behaviors are representative of private behavior and whether deviation from typical behavior is of consequence in testing hypotheses.

We used participants' ratings of the representativeness of their partner's behavior during discussions with each other to examine two basic questions. (a) Is there meaningful variance in the ratings of typicality of self-presentation? That is, will people be willing to say that their partner was not behaving typically, even if saying so may be socially undesirable? (b) If atypical behavior does occur during the interaction, does it affect the predictive validity of samples of observed behavior?

This last question is reasonable because one could obtain valid assessments from each partner of whether the other partner had acted in a nonrepresentative manner, and yet such bias might not have any effect on the predictive validity of the observed measures. Such an effect could occur if the amount of bias were not sufficient to change participants' standings in the distribution of scores on undermining (or whatever else was being observed), in which case the assessment procedure would still be able to derive valid observational data.

To address these issues, we examined the effect of typicality on the relationship of observed behavior to criterion variables. In this study, the observations were behavior samples collected from couples discussing a major stressor in their lives, job loss (e.g., Atkinson, Liem, & Liem, 1986; Caplan, Vinokur, Price, & van Ryn, 1989; Rook, Dooley, & Catalano, 1991). The criterion variables were (a) the other partner's perceptions of the general levels of supportiveness and undermining of their partner's behavior and (b) dyadic adjustment.

Method

Sample

The data were collected from 239 recently unemployed respondents and their partners, who had volunteered as part of a larger study on the impact of social support and undermining on dyadic adjustment and coping among couples facing a major stressor, job loss (Howe, Caplan, Foster, Lockshin, & McGrath, 1995).

Participants were recruited from an initial pool of 25,773 clients from five state unemployment offices in southern Maryland. To be eligible, the job seekers had to have been laid off permanently from a nonseasonal job within the last 49 days, be actively seeking reemployment of at least 20 hr per week, be married or living with someone in a permanent committed heterosexual relationship of at least 6 months' duration, have a telephone in their residence, and be able to read and speak English. Seventy-five percent of the sample were married, and 25% were living together.

Of the 239 couples (478 participants), 54.8% were European American and 45.2% were African American. Mean age was 35.6 years. The mean level of education was 13.71 years. The sample's median total family income was between $35,000 and $39,999 in 1991, the year prior to data collection. There were no significant differences between men and women in age, reported family income, or education.

Data Collection

Data were collected during an in-home interview by a trained interviewer. Each member of the couple completed several survey measures and participated in a 15-min videotaped discussion with his or her partner regarding how they were coping with the job loss. Except for the videotaped discussion, respondents were interviewed in separate rooms.

After completing nearly half of the survey measures, the job seeker and partner were each primed separately for the discussion. To do so, the interviewer asked a series of questions about how things had changed in the home and in their relationship since the job loss had occurred, and about what they wanted to see changed in their relationship.

After being prepared, the man and woman were seated in front of a video camera and told that they would have 15 min to discuss these issues. If they finished early, they were instructed to remain seated and talk about other things until the time was up. Then, the interviewer left the room.

On completion of the videotaped discussion, the job seeker and partner were again placed in separate rooms and asked to rate one another's behavior that had occurred during the discussion. After these ratings were completed, each respondent completed the rest of the survey.

Measures

All questionnaire measures used in this study were multi-item indexes with coefficients alpha of greater than .73. Table 1 presents the means, standard deviations, coefficients alpha, and intercorrelations among the measures for male and female partners separately.

Observational coding of emotional validation and invalidation during the behavioral sample. These measures were based on a behavioral coding system, the *Codebook of Marital and Family Interaction* (COMFI) by Notarius, Pelligrini, and Martin (1991). The COMFI was created specifically to assess couples engaged in a problem-solving discussion. It consists of 31 individual codes that can be collapsed into a series of six summary categories. Two of the six summary categories, emotional validation and emotional invalidation, were selected for use in the analyses because they most closely corresponded with our measures of perceived social support and undermining. The emotional validation category consists of codes such as positive feedback and primary support. The emotional invalidation category consists of codes such as guilt induction and personal criticism. Videotaped discussions were coded by assigning individual codes to thought units. Thought units are considered to be roughly equivalent to clauses.

Coders were trained until they reached an acceptable level of agreement, with criterion coding done by a separate consensus team of coders. Coders were then assigned blocks of tapes for coding. Each block included one reliability tape, randomly chosen from the complete set; coders were unaware of its status. Tapes were assigned in 28 blocks, such that approximately one tape in three was a reliability tape. Reliability tapes were coded separately by a consensus team of two or three coders, who coded separately and then met to establish consensus codes on each reliability tape. Coder ratings were compared with these consensus codes; if a coder's overall agreement level fell below a kappa of .60, the block of tapes coded by that coder was randomly reassigned to another coder for recoding in a future block. Kappa agreement statistics for acceptable blocks ranged from .60 to .93, with a mean of .69.

The amount of validating and invalidating behavior the discussants

Table 1

Intercorrelations, Means, Standard Deviations, and Coefficients Alpha Among All Variables by Gender

Variable	1	2	3	4	5	6	7	8	9
1. % of received emotional validation	—	-.42***	.10	.02	.27***	.28***	-.26***	-.25***	.33***
2. % of received emotional invalidation	-.36***	—	-.14**	.12	-.49***	-.47***	.55***	.49***	-.53***
3. Typicality of partner's support	.06	-.12	—	.42***	.12	.41***	-.12	-.28***	.05
4. Typicality of partner's undermining	-.08	.05	.37***	—	.06	-.20***	.04	.12	-.01
5. Received support (general)	.28***	-.35***	.13**	-.01	—	.64***	-.67***	-.53***	.75***
6. Received support (discussion)	.30***	-.39***	.30***	-.08	.69***	—	-.47***	-.50***	.52***
7. Received undermining (general)	-.30***	.42***	-.19***	.00	-.66***	-.52***	—	.65***	-.74***
8. Received undermining (discussion)	-.34***	.45***	-.30***	.07	-.47***	-.52***	.61***	—	-.51***
9. Marital adjustment	.25***	-.37***	.14**	.05	.63***	.49***	-.67***	-.50***	—
Men									
M	0.12	0.13	2.20	1.77	3.79	3.63	1.86	1.47	106.43
SD	0.09	0.13	0.71	0.63	0.88	1.00	0.84	0.78	20.63
α	—	—	.85	.78	.93	.89	.91	.86	.95
Women									
M	0.13	0.20	2.17	1.73	3.64	3.42	1.89	1.47	106.03
SD	0.10	0.17	0.74	0.64	0.98	1.05	0.85	0.84	21.52
α	—	—	.88	.73	.95	.89	.91	.89	.94

Note. The relationships between typicality variables are represented by phi coefficients. Data for women appear in the upper diagonal, data for men appear in the lower diagonal.
** $p < .05$. *** $p < .01$.

provided was operationalized as the percentage of codes in the emotional validation or invalidation categories, respectively, that were assigned to each person. The percentage of behavior representing each category was calculated by dividing the number of codes in each category by the total number of codes for that person. For example, if a partner emitted 300 codable behaviors and 100 of these were emotional validation codes, then the proportion of that partner's behavior coded as being supportive would be 33%. Because these estimates were based on proportional behavioral data, we tested whether arcsin transformations improved the distributional properties of the data. Such transformations made no difference in the results of any of the analyses reported here.

Typicality of social support and social undermining. The measures for these constructs were developed for this study and assessed the respondent's perceptions of how typical was his or her partner's 15 min sample of supportive and undermining behaviors. A typical item read "Compared to the usual difficult or stressful conversations between us, the amount of sympathy and understanding which my partner showed was . . ." Respondents rated the typicality of their partner's behavior on a 5-point scale ranging from 1 = *much less than usual* to 5 = *much more than usual.* The exact items are shown in the Appendix.

Principal-components analysis with varimax rotation was conducted on the eight-item measure. It revealed two factors that accounted for 63.0% of the variance in the items. Table 2 presents the factor loadings and item content. Five of the items dealing with typicality of social support loaded highly on Factor 1. Three items dealing with typicality of social undermining loaded highly on Factor 2. Item 7 was deleted from the second factor, however, because its inclusion would have significantly reduced the scale's internal reliability.

Relatively few partners were rated in the most extreme categories. For "much less" and "much more" than usual support, there were 1 and 12 persons, respectively; for "much less" and "much more" undermining, there were 28 and 3 persons, respectively. Consequently, persons in these extreme categories were combined with persons in the next most extreme categories. As a result, respondents were categorized into three groups: those showing usual, more than usual, and less than usual amounts of the observed behavior.

Perceived social support and social undermining. The measures of perceived supportive and undermining behaviors from one's partner were based on research by Caplan and colleagues sampled from other popula-

tions including adults as well as university students (Abbey et al., 1985; Vinokur et al., 1987; Vinokur, Price, & Caplan, 1996). Perceived social support and social undermining were each measured at multiple time points during the interview. Respondents were asked to indicate how supportive and undermining their partner was (a) during the interaction task as well as (b) in general. Ratings of perceived social support and undermining during the discussion were collected immediately after the completion of the interaction task. After completing several interpolated measures, respondents were asked to rate how supportive and undermining their partner was in general.

Such interpolation should minimize contamination of the second measure of support by the first one. Analyses discussed in the Results section suggest that although the postobservational and general measures of support were positively correlated, they had different predictors. Consequently, even if contamination could have occurred, it would not have been sufficient to undermine the discriminant validity of the two measures.

The measure assessing the videotaped interaction began with the phrase, "In the videotaped discussion you and your partner just had, how much did your partner . . ."; the questionnaire measure began with the phrase "How much does your partner . . . ." Respondents rated on 5-point scales the amount of support and undermining they received, which ranged from 1 = *not at all* to 5 = *a great deal.*

The indicators of *perceived social support* asked, "How much [did/does] your spouse or partner . . . [provide you with encouragement] [provide you with useful information] [say things that raise your self confidence] [listen to you when you need to talk] [show that he/she cares about you as a person] [understand the way you think and feel about things], [talk with you when you are upset] [help you understand and sort things out] [provide you with direct help] [make you feel you can rely on him/her]."

Perceived social undermining was assessed by asking "How much does your spouse or partner . . . [act in an unpleasant or angry manner towards you] [make your life difficult] [act in ways that show he/she dislikes you] [make you feel unwanted] [get on your nerves] [criticize you] [insult you even if he/she did not mean to]."

Dyadic adjustment was measured using the 32-item Dyadic Adjustment Scale (DAS) developed by Spanier (1976). This measure assesses dyadic adjustment along four different dimensions: Dyadic Cohesion,

Table 2
Factor Loadings of Typicality Items

Item no.	Description	Factor 1	Factor 2
1.	The amount of *sympathy and understanding* which my partner just showed was	.82	
2.	The amount of *care and concern* which my partner just showed toward me was	.80	
3.	The amount of *unpleasantness or negativeness* my partner just showed toward me was		.86
4.	The degree to which my partner thought *mostly about himself or herself* rather than both of us was		.85
5.	The degree to which my partner saw us as needing to work *together* as a *couple* was	.80	
6.	The degree to which my partner acted in a way that *made me feel good* about myself was	.84	
7.	The degree to which my partner *dominated the conversation* was		.54
8.	The degree to which my partner *understood* and gave me what I wanted to get out of the conversation was	.73	

Dyadic Adjustment, Dyadic Consensus, and Affectional Expression. Only the total DAS score was used in this study.

Methods of Analysis: A Word About the Inherent Nonindependence of Couples' Data

The observational measures as well as the criterion variables of perceived social support and dyadic adjustment derive their meaning by being reactions of each partner to the other. Accordingly, the findings for one partner are not independent of those for the other, either psychometrically or in nature. Preliminary analyses showed that when we attempted to remove variance shared by the couple, such as by removing the mean of the couple's dependent variable (Pedhazur, 1982), virtually all (99%) of the variance in the dependent variable would be removed. As a result, any residual effects of gender would be those that were irrelevant to the observed social interaction.

Rather than take such draconian steps, we present the findings by gender with the understanding that (a) each gender group's findings on their own represent a set of independent observations and that (b) these subgroups should not be viewed as independent replications of each other. There is little evidence of gender effects, but where there might be, we prefer to view the evidence in terms of effect size rather than significance and as having occurred within the context of a couple relationship.

Where this article reports a finding and draws from the data for both the men and women (e.g., "the correlations were in the mid-30s to low 40s"), the use of all of the coefficients is made to better characterize the effect size of the findings as a whole and not to characterize the statistical significance. With regard to significance, and as the most conservative estimate, the degrees of freedom remain that of the number of couples (239) rather than that of the total of all observations for men and women (478).

Results

This section begins with an examination of the criterion validity of the observational measures. Then we look at the pattern of rated typicality of behavior and consider the likelihood that the ratings of bias are, themselves, subject to distortion. We conclude by examining whether such ratings can identify subgroups of persons whose observational measures show poor

criterion validity. Unless noted otherwise, statistical coefficients presented in the text are statistically significant.

Validity of the Observational Measures

As seen in Table 1, the correlations of observed validation and invalidation with other measures suggest that the observational data are valid indicators of the general quality of social relationships in the couple and replicate findings from previous research: (a) High levels of emotional validation were most likely to be observed in relationships where the recipients reported generally high social support, low social undermining, and high dyadic adjustment. (b) High levels of emotional invalidation were associated at about the same magnitude with low social support, high social undermining, and low dyadic adjustment. (c) Observed validation and invalidation were inversely related, as were recipients' postobservation ratings of support and undermining. (d) Participants' ratings of their partners' behavior during the interaction agreed with codings by independent raters.

Overall, these findings suggest that the observational measures are reasonably valid. The question now becomes whether variation in rated typicality modifies the relationship between observed behavior and general ratings of the couple's relationship.

Distribution of Atypical Behavior in the Sample

As shown in Table 3, partners judged each other as behaving typically with regard to supportive and undermining behavior from 42% and 55% of the time. Where partners judged one another to be acting atypically, partners were about twice as likely to rate each other as being more rather than less supportive than usual and more than three times as likely to rate their partners as being less rather than more undermining than usual. These findings provide evidence that there is substantial variance in ratings of typicality and that the variance away from typicality is in directions that would be predicted by previous theory and research on the importance of social approval in self-presenta-

Table 3
Frequency Distribution of Typicality Ratings by Gender

Typicality of . . .	Amount		
	Less than usual	Same as usual	More than usual
Typicality rating of woman by man			
Social support			
Frequency	40	107	88
Row %	17.0	45.5	37.4
Social undermining			
Frequency	77	126	24
Row %	33.9	55.5	10.6
Typicality rating of man by woman			
Social support			
Frequency	48	99	88
Row %	20.4	42.1	37.4
Social undermining			
Frequency	86	117	24
Row %	37.9	51.5	10.6

tion (e.g., Allport, 1955; Crowne & Marlowe, 1964; Leary & Kowalski, 1990).[1] Men and women did not differ significantly in the distribution of their ratings of their partner's typicality of social support, $\chi^2(1, N = 470) = .55$, *ns* and undermining, $\chi^2(1, N = 454) = .72$, *ns*.

Are the Measures of Typicality Subject to Bias?

If the measures of typicality are themselves free of bias, we might expect that they would be unrelated to preexisting properties of each rater, particularly the rater's perceptions of relationship quality and well-being. Such a lack of relationship is important because the measures of preexisting relationship quality form the criterion measures when determining whether typicality moderates criterion validity of the observational data. To test these expectations and because the typicality measures are not necessarily ordinal scales, we examined both potential linear and nonlinear associations between measures of typicality and the rater's perception of relationship quality and well-being. In general, the pattern of findings suggests that the ratings of typicality were unrelated to preexisting attitudes of the rater regarding the quality of the social relationship and, as such, were not subject to biases of substantive interest. Only the linear effects are presented in Table 1. Most associations were nonsignificant, and those few that were significant were weak.

The only relationships of note with typicality deal with its association with measures of support and undermining expressed during the behavior sample itself. The more the recipient perceived the partner as acting more supportive than usual, the more the recipient viewed that behavior as supportive per se and the less the recipient viewed the behavior as undermining per se (*r* ranged in absolute value from .28 to .41). These associations make sense in that it is likely that the more the recipient perceived that the partner was showing high levels of social support, the more likely that support would be seen as atypically high rather than atypically low. Both level of support

and level of perceived typicality might reasonably influence each other during the interaction sample. Typicality of undermining, however, failed to show comparable associations with perceived support and undermining during the interaction (*r*s, which were mostly about 0, ranged in absolute value from .07 to .20). The variances of typicality of social support and undermining were similar, so differences in variance could not have accounted for this pattern of results. One hypothesis suggested by these results is that people are generally optimists who want to believe that increases in positive acts toward them are sincere. Alternatively, decreases in negative acts may be seen by partners as attempts to merely avoid conflict rather than as an attempt to solve it.

Biased Self-Presentation: Does It Undermine the Criterion Validity of Observed Behaviors?

In these analyses, we examined the role of ratings of typicality as a moderator of the expected links between participants' observed behaviors and three criterion variables: (a) their partner's more general ratings of perceived support and (b) undermining in the relationship, and (c) dyadic adjustment. It was predicted that the slope of the relationship between each participant's observed behavior and his or her partner's report of support, undermining, and dyadic adjustment would be weakest for participants rated as behaving either more supportively or less undermining than usual.

First, to minimize multicollinearity among predictors, the data were centered using procedures described by Aiken and West (1991). Next, a stepwise moderated multiple regression was used to test for moderator effects using an equation of the general form $y = [intercept] + b_1 X + b_2 Z + b_3 XZ$, where *b* is the estimate of the particular slope, *y* represents the criterion variable, *X* represents an index of the observed behavior; *Z* represents the hypothesized moderating variable, typicality group; and *XZ* represent the moderating effects of typicality on observed partner behavior.

After the interaction term was entered into the equation, the change in R^2 was tested for significance. Moderator effects of typicality ratings were tested separately for each index of typicality: social support and social undermining.

There were six sets of tests for interaction effects involving two moderators (typicality of social support and of undermining) and three criterion variables (social support in general, social undermining in general, and dyadic adjustment). Table 4 summarizes the significant findings.

Although typicality of support and undermining were positively and moderately correlated, their effects were analyzed separately to determine whether it would be important to distinguish between them. The effects of the moderators were examined with regard to both their commensurate measures of observed behavior (e.g., typicality of social support moderating the effects of emotional validation, typicality of undermining

---

[1] When partners were rated as displaying more undermining behavior than usual, such ratings may have reflected the rater's need to present his or her partner in a desirable light. Such raters in effect may have been saying to the researcher "my partner is usually not this undermining, so excuse my partner." The design did not permit us to determine the validity of such interpretations.

Table 4

Summary of Hierarchical Regression Analysis for Effects of Typicality of Social Support on the Validity of Observed Behavior

Moderator	Dependent variable	Global F	p	ΔR^2	Simple slopes			t tests	t
					<usual (A)	usual (B)	>usual (C)		
				Woman's emotional validation					
Typicality of social support	Man's dyadic adjustment	3.96	.02	.030	137.41	69.71	19.84	A vs. B A vs. C B vs. C	−2.65***
	Man's recv'd support	1.97	.14	.015	4.77	3.23	1.43	A vs. B A vs. C B vs. C	−1.74*
	Man's recv'd undermining	3.64	.03	.029	−31.53	−5.10	−1.82	A vs. B A vs. C B vs. C	2.63***
				Man's emotional validation					
Typicality of social support	Woman's dyadic adjustment	2.37	.09	.020	119.57	86.38	34.99	A vs. B A vs. C B vs. C	−1.76* −1.76*
	Woman's recv'd support	5.12	.01	.039	8.86	3.42	1.98	A vs. B A vs. C B vs. C	−2.42** −3.19***
	Woman's recv'd undermining	1.97	.15	.016	−5.37	−2.50	−1.97	A vs. B A vs. C B vs. C	1.98**

Note. recv'd = received.
* $p < .10$.   ** $p < .05$.   *** $p < .01$.

moderating emotional invalidation) and their noncommensurate measures (e.g., typicality of social support moderating effects of emotional invalidation). As expected, the findings for the commensurate measures were greater in number and yielded more consistent findings. Consequently, the results focus on the commensurate analyses. Given the exploratory nature of the study and the weak power of statistical tests in detecting interactions, we paid attention to the pattern of the findings as well as their statistical significance.

Biased Self-Presentation Was More Likely for Validating Than Invalidating Behavior

In 11 of 12 analyses involving the combinations of criterion variables and measures of typicality, the validity coefficients of the simple slopes were weakest, as predicted, for persons who behaved more supportively or less undermining than usual. As shown in Table 2, these were the most common expressions for atypical behavior.

Overall, however, the moderating effects of typicality on criterion validity of emotional validation behavior were more frequent (6 cases vs. 2 cases where slopes differed as a function of typicality) and interpretable than those for effects on the criterion validity of emotional invalidation behavior. Where moderating effects of typicality on invalidating behavior did occur, they were near or at the margins of significance set for the interaction effects ($p < .10$). Consequently, we focused on how emotional validating behavior was affected. Table 4 summarizes those results.

Where there was significant ($p < .05$) evidence of moderating

effects of typicality, those effects accounted for approximately 2 to 4 additional percentage points of variance in the criterion variables. For both the male and female subgroups, evidence of losses in criterion validity due to atypical behavior occurred more often during the observation of emotionally validating than invalidating behaviors.

Most Valid Data Came From Persons Rated as Behaving Less Supportive Than Usual

Contrary to hypothesis, the group with the highest criterion validity showed less than usual amounts of supportive behavior (average absolute magnitude of the slope = 51.25). The group expected to show the highest validity coefficients, persons rated as "typical," came in second, averaging only 28.39; the group showing more than typical supportiveness averaged 10.34. Consequently, the most striking difference in validity coefficients generally occurred between persons behaving more supportively and less supportively than usual. Figure 1 illustrates these differences with graphs of the relative heights for four of these sets of slopes.

For the two marginally significant findings dealing with the validity of emotionally invalidating behavior (not shown), the picture was less clear. Although the group showing less than usual invalidation had the lowest validity coefficient, as predicted, and the usual or "typical" group had the highest validity coefficient, that coefficient was not significantly different from the group that showed more invalidating behavior than usual.

The criterion measures of social support and social undermining, as well as dyadic adjustment, were significantly intercorre-
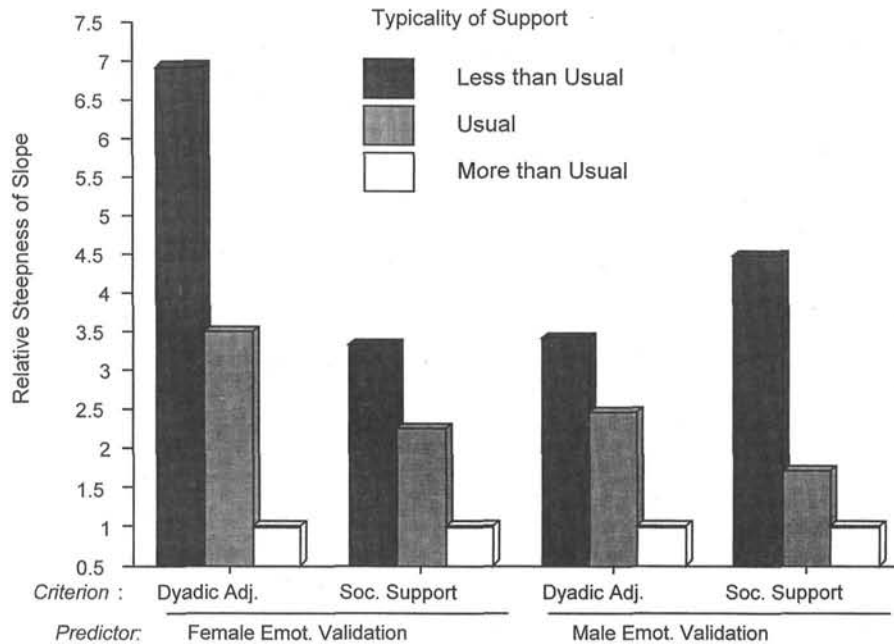
Figure 1. Impact (regression slope steepness) of emotionally validating behavior on selected criteria as a function of typicality of behavior. Dyadic Adj = dyadic adjustment; Soc. Support = social support; Emot. Validation = emotional validation.

lated, so the findings for each criterion are not necessarily independent. Consequently, for emotionally validating behavior, social undermining was as likely as social support to serve as a criterion variable, and no one criterion was superior as an indicator of the moderator effects.

Discussion

This study explored whether partners providing observational samples of their behavior might provide valid reports of the representativeness of each other's behavior during those samples. We could find only one other similar study (Warren & Gilner, 1978), which was conducted nearly 20 years ago. It used limited measures of indeterminate reliability and reported no significant findings.

Although the current study's findings replicate across men and women and involve couples across a wide range of social status levels, this study also has its limits. The sample consists of couples dealing with a particular recent life event, job loss, and only with two particular characteristics of couple interaction, validation and invalidation. It is unknown whether similar results would be generated when studying other life events or settings and other characteristics of interaction, such as agenda setting and inhibition. Nor is it known whether similar results might be generated by examining the criterion validity of sequences of exchange between people (e.g., Gottman, 1979) rather than mean levels of exchange. Given the paucity of research, it is important that the results be replicated. In discussing the findings, we offer some suggestions on how replication and extension should proceed.

Can People Make Meaningful Judgments About the Typicality of One Another's Behavior?

When a sample of adults was videotaped in their homes while having a discussion about a serious topic in their lives, they generated significant, nonrandom variation in how they rated the typicality of one another's supportive and undermining behavior. Nearly one half the participants were rated by their partners as behaving typically in terms of socially supportive and undermining behavior. Among the remainder, participants were more likely to be rated as behaving in socially desirable (i.e., providing more support and less undermining than usual) than undesirable directions. This pattern of results was the same for both female and male partners. The direction of the results is consistent with theory and research regarding the tendency of people to present themselves in a socially desirable manner that enhances their public reputation (e.g., Allport, 1955; Doby & Caplan, 1995; Leary & Kowalski, 1990). The findings suggest that such ratings were not influenced or biased by preexisting attitudes regarding the quality of the relationship that was to be predicted by the observational samples.

These ratings of typicality were unrelated to the observational measures of invalidation and validation, were moderately related to the rating partner's judgments of how supportive the other partner had been during the interaction, and were weakly or unrelated to the rating partner's judgments of how supportive the other partner was in general in the relationship. The low or nonsignificant associations of typicality ratings with the measures of perceived general levels of support and undermining point to the appropriateness of using the more general measures

as independent criteria against which to judge the validity of the observational measures.

Although the study made use of previously validated measures of relationship quality and well-being that showed interrelations similar to those found in other studies, the approach to measuring typicality was novel and exploratory. The index of typicality of undermining was based on only two items, whereas that of support was based on five items. Although the reliability of both scales is adequate, the construct and domain validity of the measure of typicality of undermining could be increased by building more items based on the detailed codes in behavior coding manuals used in this study and elsewhere (e.g., Gottman, Notarius, Gonso, & Markman, 1976).

Hypothesized Mechanisms and Implications for Future Research

Why might the criterion validity of emotional validation be particularly susceptible to the effects of atypical self-presentation? Typicality was more likely to affect the criterion validity of emotional validation compared with emotional invalidation. This effect could occur if people being atypically invalidating, for reasons explored later, made inconsequentially small shifts away from their usual behavior. With only small shifts, the rank ordering of these persons might not change from what it was outside of the observational sample, and the expected association of invalidating behavior with criterion measures would not be degraded.

Smaller shifts could occur if it was more difficult for persons to alter their base rate of invalidating compared with validating behavior. Previous research supports this finding. This pattern of results suggests that it may be easier for people in public situations to increase positive behavior rather than suppress negative or interpersonally critical behavior. This would be the case if emotionally invalidating behavior is more self-sustaining during the couple's interaction, as Howe and Reiss (1993) have suggested, and accordingly resistant to attempts to alter it. This interpretation is supported by the finding that couples embroiled in cross-complaining cycles and patterns of negative reciprocity have difficulty attenuating such behavior (Gottman, 1979).

A direct test of this and several other hypotheses discussed later might be possible using a within-persons, repeated-measures experiment. Partners might be exposed to six conditions involving instructions to behave at three levels of typicality— less than typical, typical, and more than typical—with regard to two dimensions of behavior—invalidating and validating. It would then be possible to determine whether the shift from typical behavior varied as a joint function of typicality and emotional validation or invalidation behavior.

Why might persons being less supportive than usual generate the most psychometrically valid data? The joint effects of two mechanisms might explain the unexpected finding that persons behaving in an apparently less socially desirable direction than usual had the highest validity coefficients. One hypothesized mechanism involves perceptual novelty (Ericcson & Simon, 1980). If one partner begins acting in a novel or atypical fashion, the other partner's attention to that behavior may be increased. Following Ericcson and Simon's (1980) hypothesis, increases in attention should increase the accuracy with which

the observing partner makes judgments regarding typicality. If so, there will be fewer false hits in assigning partners to either the more-than- or less-than-typical groups than to the typical group. Assignments to the typical category would involve more false hits because such judgments are likely to be made by raters whose attention to the partner's behavior is relatively low.

The repeated-measures experiment described earlier could test this effect by examining changes in within-observer accuracy as the observed partner. After viewing the acting partner's behavior in the conditions described earlier, the observing partner might be asked to "think-aloud" to an interviewer while completing the postinteraction ratings of typicality (Sudman, Bradburn, & Schwarz, 1996). Such techniques could permit the investigator to be a party to the elements entering the rater's decision regarding level of typicality. Physiological measures of the observing partner's orienting response might be assessed (Frankenhaeuser, 1976) as a second channel for assessing hypothesized effects of novelty on rater attention.

The "rater attention" hypothesis would partially explain the distribution of validity coefficients in Figure 1, the reason why the "typical" group would show only moderately high rather than very high levels of criterion validity. To explain why the atypical groups might differ from each other in criterion validity, we turn to hypotheses regarding whether different processes drive more-than-typical compared with less-than-typical amounts of emotionally validating behavior.

The generation of more-than-typical validating behavior might be driven by the need to create a positive impression on the researcher (Leary & Kowalski, 1990). Although the need might be common among persons acting more supportively than usual, there could be significant individual differences within that group in the ability to generate such behavior (e.g., Riggio, Watring, & Throckmorton, 1993). If so, such differences could undermine the overall validity of behavioral samples produced by the more-than-typical group. This explanation is consistent with the original hypothesis of the study regarding the psychometrically invalidating effects of acting atypically.

The generation of less-than-typical validating behavior may be due to two mechanisms. Both shyness (Shepperd & Arkin, 1990) and desire to cooperate with the investigator (Orne, 1969) by being more "honest" and "letting it all hang out" might reduce the frequency of validating behavior. For these mechanisms to work, one would need to hypothesize that within both shy and cooperative persons there might be few individual differences in the ability to suppress positive behavior toward one's partner. A lack of individual differences could occur if there are fewer occasions to be less rather than more nice to one's partner in public situations. With few opportunities to practice, most people might be similar in their ability to suppress the supportiveness they usually show during problem-solving discussions with their partner. This lack of differences would leave the criterion validity of observed behavior intact for the less-than-typical group.

The "honest" subset of behaviors could additionally represent the set of verbal and nonverbal cues that the observing partner is unconsciously more sensitive to outside of the observational situation. That process could also generate high validity coefficients for the "honest" group.

The experimental paradigm described earlier could be ex-

tended to test these hypotheses. The measurement of shyness and of participant beliefs about what types of behaviors might be desirable during the observation sample (e.g., Rosenthal & Rosnow, 1991, p. 117) could be entered into the analyses to determine whether they account for variance in the criterion validity of the results.

In Sum

Over the years, couples participating in this and other studies have willingly provided observational samples of their interaction. Findings from many of these studies demonstrate that such observations can predict relationship quality and the long-range stability of the relationship. Reviews of the literature, however, show that many samples, particularly those using observational data, are considerably smaller than the 239 couples in this study and are likely to have difficulty detecting interaction effects of even moderate size (see reviews by Karney & Bradbury, 1995; Sullaway & Christensen, 1983). Given the low power of the typical study, there is still a continuing need to determine and model the effects of sources of invalidity that reduce the power of such studies. Should the significant moderating effects of self-presentation bias replicate and be extended, such an approach may provide a useful tool in estimating the true effects of behavior on relationship quality. Discovering the best way to generate such estimates remains a challenge. Asking the partner's help in generating that estimate could be a promising line of inquiry.

References

Abbey, A., Abramis, D. J., & Caplan, R. D. (1985). Effects of different sources of social support and social conflict on emotional well-being. *Basic and Applied Social Psychology, 6,* 111–129.

Adler, A. (1930). Individual psychology. In C. Murchinson (Ed.), *Psychologies of 1930* (pp. 395–405). Worcester, MA: Clark University Press.

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions.* Newbury Park, CA: Sage.

Allport, G. (1955). *Becoming.* New Haven, CT: Yale University Press.

Atkinson, T., Liem, R., & Liem, J. (1986). The social costs of unemployment: Implications for social support. *Journal of Health and Social Behavior, 27,* 317–331.

Caplan, R. D., Vinokur, A. D., Price, R. H., & van Ryn, M. (1989). Job seeking, reemployment, and mental health: A randomized field experiment in coping with job loss. *Journal of Applied Psychology, 74,* 759–769.

Cobb, S. (1976). Social support as a moderator of life stress. *Psychosomatic Medicine, 38,* 300–314.

Cohen, S., & Wills, T. H. (1985). Stress, social support and the buffering hypothesis. *Psychological Bulletin, 98,* 310–357.

Coyne, J. C., & Bolger, N. (1990). Doing without social support as an explanatory concept. *Journal of Social and Clinical Psychology, 9*(1), 148–158.

Coyne, J. C., & DeLongis, A. (1986). Going beyond social support: The role of social relationships in adaptation. *Journal of Consulting and Clinical Psychology, 54,* 454–460.

Crowne, D. P., & Marlowe, D. (1964). *The approval motive.* New York: Wiley.

Doby, V. J., & Caplan, R. D. (1995). Organizational stress as a threat to reputation: Effects on anxiety at work and at home. *Academy of Management Journal, 38,* 1105–1123.

Ericcson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review, 87,* 215–251.

Frankenhaeuser, M. (1976). The role of peripheral catecholamines in adaptation to understimulation and overstimulation. In G. Serban (Ed.), *Psychopathology of human adaptation* (pp. 173–191). New York: Plenum Press.

Gerin, W., Milner, D., Chawla, S., & Pickering, T. G. (1995). Social support as a moderator of cardiovascular reactivity in women: A test of the direct effects and buffering hypotheses. *Psychosomatic Medicine, 57,* 16–22.

Gottman, J. M. (1979). *Marital interaction: Experimental investigations.* New York: Academic Press.

Gottman, J. M., Notarius, C. I., Gonso, J., & Markman, H. J. (1976). *A couple's guide to communication.* Champaign, IL: Research Press.

Howe, G. W., Caplan, R. D., Foster, D., Lockshin, M., & McGrath, C. (1995). A research strategy for developing preventive interventions. In G. Keita & S. Sauter (Eds.), *Stress in the 90s* (pp. 139–158). Washington, DC: American Psychological Association.

Howe, G. W., & Reiss, D. (1993). Simulations and experimentation in family research. In P. G. Boss, W. J. Doherty, R. LaRossa, W. R. Schumm, & S. K. Steinmetz (Eds.), *Sourcebook of family theories and methods: A contextual approach.* New York: Plenum Press.

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.

Kashy, D. A., & Snyder, D. K. (1995). Measurement and data analytic issues in couples research. *Psychological Assessment, 7,* 338–348.

Leary, M. R., & Kowalski, R. M. (1990). Impression management: A literature review and two component model. *Psychological Bulletin, 107,* 34–47.

Markman, H. J., Renick, M. J., Floyd, F. J., & Stanley, S. M. (1993). Preventing marital distress through communication and conflict management training: A 4- and 5-year follow-up. Special Section: Couples and couple therapy. *Journal of Consulting and Clinical Psychology, 61,* 70–77.

Notarius, C. I., Pelligrini, D. S., & Martin, L. A. (1991). *Codebook of Marital and Family Interaction (COMFI).* Unpublished manuscript.

Orne, M. T. (1969). Demand characteristics and the concept of quasi-controls. In R. Rosenthal & R. L. Rosnow (Eds.), *Artifact in behavioral research* (pp. 143–179). New York: Academic Press.

Pedhazur, E. J. (1982). *Multiple regression in behavioral research: Explanation and prediction.* New York: Holt, Rinehart, and Winston.

Rhodes, J. E., & Woods, M. (1995). Comfort and conflict in the relationships of pregnant, minority adolescents: Social support as a moderator of social strain. *Journal of Community Psychology, 23,* 74–84.

Riggio, R. E., Watring, K. P., & Throckmorton, B. (1993). Social skills, social support, and social adjustment. *Personality & Individual Differences, 15*(3), 275–280.

Rook, K. S. (1984). The negative side of social interaction: Impact on psychological well-being. *Journal of Personality and Social Psychology, 46,* 1097–1108.

Rook, K., Dooley, D., & Catalano, R. (1991). Age differences in workers' efforts to cope with economic distress. In J. Eckenrode (Ed.), *The social context of coping* (pp. 79–105). New York: Plenum Press.

Rosenberg, M. J. (1979). *Conceiving the self.* New York: Basic Books.

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

Sarason, B. R., Sarason, I. G., & Pierce, G. R. (Eds.). (1990). Traditional views of social support and their impact on assessment. *Social support: An interactional view* (1st ed., pp. 9–25). New York: Wiley.

Shepperd, J. A., & Arkin, R. M. (1990). Shyness and self-presentation. In W. R. Crozier (Ed.), *Shyness and embarrassment: Perspectives from social psychology* (pp. 286–314). New York: Cambridge University Press.

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

Sudman, S., Bradburn, N. M., & Schwarz, N. (1996). *Thinking about answers: The application of cognitive processes to survey methodology.* San Francisco: Jossey-Bass.

Sullaway, M., & Christensen, A. (1983). Couples and families as participant observers of their interaction. In J. P. Vincent (Ed.), *Advances in family intervention, assessment, and theory* (Vol. 3, pp. 119–160). Greenwich, CT: JAI Press.

Vinokur, A. D., Price, R. H., & Caplan, R. D. (1996). Hard times and hurtful partners: How financial strain and social undermining affect relationship satisfaction and depression of unemployed persons and their spouses. *Journal of Personality and Social Psychology, 71,* 166–179.

Vinokur, A., Schul, Y., & Caplan, R. D. (1987). Determinants of perceived social support: Interpersonal transactions, personal outlook, and transient affective states. *Journal of Personality and Social Psychology, 32,* 329–337.

Warren, N. J., & Gilner, F. H. (1978). Measurement of positive assertiveness behaviors: The behavioral test of tenderness expression. *Behavior Therapy, 9,* 178–184.

Wethington, E., & Kessler, R. C. (1986). Perceived support, received support, and adjustment to stressful life events. *Journal of Health and Social Behavior, 27,* 78–89.

Appendix

Measure of Typicality

We need your help in making some comparisons between the conversation you just had and a usual discussion you might have with your partner about difficult or stressful subjects.

Compared to the usual difficult or stressful conversation(s) between us,

	Much less than usual	Somewhat less than usual	About usual	Somewhat more than usual	Much more than usual
1. The amount of sympathy and understanding which my partner just showed was	1	2	3	4	5
2. The amount of *care and concern* which my partner just showed toward me was	1	2	3	4	5
3. The amount of *unpleasantness or negativeness* my partner just showed toward me was	1	2	3	4	5
4. The degree to which my partner thought *mostly about himself or herself* rather than both of us was	1	2	3	4	5
5. The degree to which my partner saw us as needing to work *together* as a *couple* was	1	2	3	4	5
6. The degree to which my partner acted in a way that made me feel good about myself was	1	2	3	4	5
7. The degree to which my partner *dominated the conversation* was	1	2	3	4	5
8. The degree to which my partner *understood* and gave me what I wanted to get out of the conversation was	1	2	3	4	5