Published in final edited form as:
J Marriage Fam. 2007 February ; 69(1): 86–91.

Generalizability Issues in Observational Studies of Couples: Sample Characteristics and Task Design

Hyoun K. Kim and
Oregon Social Learning Center, 10 Shelton McMurphey Blvd. Eugene, OR 97401, (541) 485-2711, (541) 485-7087, hyounk@oslc.org

Deborah M. Capaldi
Oregon Social Learning Center, 10 Shelton McMurpphey Blvd. Eugene OR 97401, (541) 485-2711, (541) 485-7087, deborahc@oslc.org

Generalizability Issues in Observational Studies of Couples: Sample Characteristics and Task Design

In response to our study regarding the generalizability of the affective process models that were proposed by Gottman, Coan, Carrère, and Swanson (1998) to predict relationship status and satisfaction in the Oregon Youth Study couples sample (Kim, Capaldi, & Crosby), Coan and Gottman (this issue) question whether tests of generalizability of their findings are an appropriate scientific undertaking, particularly when the Oregon Youth Study sample had characteristics that were so different from those of the Gottman et al. sample. They argue that the Oregon Youth Study sample is so specialized as to have limited generalizability to other U.S. couples. Coan and Gottman further argue that their affective process models were not meant to be generalizable to other samples, which included cohabiting couples and couples who are predominantly from low-income and at-risk backgrounds.

We believe that the majority of scholars in this field (e.g., Heyman, 2001), as well as other fields (e.g., Mihalic, Irwin, Elliott, Fagan, & Hansen, 2001), consider tests of generalizability critical to move the field forward in credibility and applicability to prevention work. Thus, the degree to which affective process functions similarly (or dissimilarly) for couples from heterogeneous backgrounds is an important question, so that any limits to generalizability may be clearly understood. As Gottman's work has drawn considerably more attention from the media, as well as clinicians and the general public, than is typical in the field, testing generalizability is particularly important. We believe our research contributes to the understanding of factors involved in functional and dysfunctional communication in couples by extending Gottman's frontier work on dynamic affective processes to couples from at-risk, low socioeconomic status (SES) backgrounds.

Below we address several issues that were raised by Coan and Gottman in their critique (this issue). In particular we present evidence (a) of the relatively high relevance of the Oregon Youth Study couples sample to a large segment of U.S. couples, (b) that cohabiting couples within the Oregon Youth Study couples sample did not account for the differences in findings between the Gottman et al. and Kim et al. studies, and (c) of the importance of examining process during discussion of both men's and women's problem issues.

Representativeness of At-Risk Sample

A major argument by Coan and Gottman (this issue) is that the Oregon Youth Study sample was selected to represent a relatively rare population, and that it likely represented the more aggressive tail of the Newlywed Study distribution. In contrast, they argue that the sample utilized in Gottman et al. (1998) was representative of the greater Seattle area. As was pointed

out by Heyman and Hunt, however, the Newlywed sample, which responded to advertising and thus was not recruited according to a specific sampling frame, was selected to match the racial and ethnic demographics of the metropolitan Seattle area only (Carrère, Buehlman, Gottman, Coan, & Ruckstuhl, 2000; Hawkins, Carrère, & Gottman, 2002). The mean education level reported for both husbands and wives was a 4-year college degree, which is relatively high. The Oregon Youth Study sample original recruitment involved *all* fourth-grade boys from schools in a medium-sized metropolitan area. Schools were selected on the criterion of having a relatively high incidence of delinquency in the neighborhoods *for that area*. As indicated by Heyman and Hunt (this issue), Eugene/Springfield in Oregon is an area with a relatively low level of violent crime (Cooper, 2006) and very little organized gang crime. It is not a high-crime area, such as an inner-city area. The sample was potentially at risk by virtue of living in higher-crime areas, which differs considerably from a high-risk sample with all participants showing elevated conduct problems at the time of recruitment. Many of the men in the Oregon Youth Study sample did not show conduct problems as youth and are well-adjusted adults. Thus, the Oregon Youth Study sample does not represent the aggressive tail of the distribution only and is thus not substantively limited by restriction of range issues on the behavioral variables.

The major effect of the neighborhood selection was that the sample was predominantly lower and working class – as is the U. S. as a whole. According to a recent study conducted by the Commonwealth Fund, approximately 35 million adults between the ages of 18 and 64 in the U. S. earn less than $20,000 a year (Budetti, Duchon, Schoen, & Shikles, 1999). Our sample was selected by design as being representative of these neighborhoods in the Eugene/Springfield area, but the evidence is that findings with this sample generalize to a broader U.S. population with similar characteristics. Our previous studies have also indicated that, although levels of antisocial behavior for our sample are somewhat elevated compared with those of other community samples, several indicators show comparable levels to those of other studies (e.g., Capaldi, Crosby, & Stoolmiler, 1996). For example, the prevalence rates for aggression toward a partner in the Oregon Youth Study of 36% for women and 31% for men at approximately age 18 years are very comparable to those of other studies with large samples. Across two U.S. studies with data collection in the mid 1980s, the National Family Violence Survey (Fagan & Browne, 1994), the National Youth Survey (Elliott, Huizinga, & Morse, 1985), as well as the Dunedin sample (Moffitt & Caspi, 1999), perpetration rates ranged from about 36% to 50% for women and 25% to 40% for men (Moffitt & Caspi). Thus, Oregon Youth Study couples are not substantially more aggressive than nationally representative samples. In addition, the Oregon Youth Study experienced a high participation rate initially (74%) and high retention rates (94%) over more than two decades. Given that families with higher-risk characteristics are less likely to participate in studies initially and more likely to drop out of studies over time (Navratil, Green, Loeber, & Lahey, 1994), our work on predominantly lower- and working-class families contributes to the field by representing a very large and understudied segment of the U.S. population.

Married Versus Cohabiting Couples

Coan and Gottman (this issue) also argue that our findings are not comparable to Gottman et al. (1998) because the Oregon Youth Study sample included cohabiting couples. We would argue that, in general, if a study is to be relevant to experiences of young couples in recent cohorts, cohabiting couples should be included, because of the rise in the prevalence of cohabitation (Manning & Smock, 2002; Seltzer, 2001; Smock, 2000). Studies have consistently indicated that married couples are different from cohabiting couples in various aspects including lower relationship satisfaction, lower interpersonal commitment, poorer communication, and higher conflict (e.g., Axinn & Barber, 1997; Cohan & Kleinbaum, 2002; Stanley, Whitton, & Markman, 2004; Thomson & Colella, 1992), and yet specific

interaction patterns that are unique to cohabiting couples, especially affective processes, have not been systematically investigated to date. A few studies on communication have indicated that cohabiting couples or married couples with cohabiting experience tended to show poorer problem-solving skills (e.g., Cohan & Kleinbaum, 2002), but the question of why cohabitation experience may be related to poorer communication remains unanswered. Cohan and Kleinbaum have suggested that it may be that the uncertain nature of the cohabiting relationship results in less commitment to the relationship and thus less motivation to develop conflict resolution and support skills. As indicated by Heyman and Hunt (this issue), however, there is not enough evidence for us to presume that the affective processes proposed by Gottman et al. (1998) would function differently for cohabiting couples, or that their predictive validity to relationship outcomes would also differ compared to that for married couples.

To address the issues of possible effects for cohabiting couples specific to our findings, we examined the Oregon Youth Study couples data to investigate whether there were any differences in mean levels of the affective process models between the married and cohabiting couples as well as between the stably married and stably cohabiting couples. As shown in Table 1, none of the affect constructs (including the level of men's negative start up) showed significant differences in mean levels between the married and cohabiting couples, nor between the stably married and stably cohabiting couples. This pattern of findings suggests that even though the married couples tended to have longer relationships ($t = -2.49$, $p = .02$), the affect processes might be similar between the married and cohabiting group in the Oregon Youth Study couples sample.

Next, we examined whether those couples who separated were substantially more aggressive than those who stayed together, as implied by Coan and Gottman (this issue). It was found that only women from the separated couples tended to show higher levels of physical aggression, compared to women from the stable couples ($t = -3.15$, $p < .01$). There was no significant difference in physical aggression between the groups for the men, or in psychological aggression between the two groups. In addition, the stably married and stably cohabiting groups did not differ from each other in aggression. These findings suggest that in the Oregon Youth Study couples sample, the cohabiting couples were very similar to the married couples in terms of aggressive behavior to each other and also for affective processes.

Cohabitation is now relatively common, especially among lower SES men and women. Using the National Longitudinal Study of the High School Class of 1982 (NLS-72), Clarkberg (1999) reported that approximately 85% of the respondents had experienced cohabitation during the 14 years of the study period. Clarkberg (1999) also found that men and women are more likely to cohabit than to marry when they are economically unstable, suggesting the cohabitation is an alternative for those who are in romantic relationships but who lack the economic well-being or occupational stability required for marriage. As such, cohabitation is more common among those with less education and limited economic resources, and for these people cohabitation is an important setting in which couples have children (Seltzer, 2001). Once they have children, their cohabiting union becomes more like a marriage (Seltzer, 2001). Some Oregon Youth Study couples had long-term cohabiting relationships, and had children. At the Time 5 assessment, the average length of the cohabiting relationships was over 3 years. For many Oregon Youth Study cohabiting couples, therefore, cohabitation is a *de facto* marriage and the couples could be very similar to married couples in many respects, such as interaction patterns and aggression. Thus, it may be problematic to assume that cohabiting couples are substantially different from married couples, or to exclude them from study. Many of the studies on cohabiting couples are on the basis of couples from middle-class backgrounds (e.g., Cohan & Kleinbaum, 2002; Dush, Cohan, & Amato, 2003; Kline et al., 2004), especially with a high education level (e.g., both husbands and wives have an average of 16 years of

education in Cohan and Kleinbaum's study). Cohabiting relationships of couples with at-risk and low SES backgrounds definitely warrant further investigation.

Men's Versus Women's Problem Issues

Coan and Gottman (this issue, p. x) argue that "Women typically start most marital conflict discussions in laboratories that use observational methods" and that our approach to collect both men and women's issues creates "a threat to ecological validity." They argue that the primary selection criteria for areas for disagreement in their sample were duration and severity and that our approach overrepresented less severe issues by requiring separate men's and women's issue (see this issue). How to set up the interaction tasks and whether having separate men's and women's issue sessions violates ecological validity would require further empirical tests regarding couples' conflictual patterns. As pointed out by Heyman and Hunt (this issue), many people in the field now acknowledge that couples' interaction patterns vary depending on the design of the discussion tasks (e.g., men's vs. women's issue, positive versus negative discussion topic). It may be that even though men may raise issues less often, the issues they do raise are highly salient to men, at least, and possibly to the long-term health of the relationship (Ball, Cowan, & Cowan, 1995). In fact, Ball et al. found that even though women tended to raise issues first in the early phase of the discussion, men controlled the content and emotional depth of the later discussion phases and, consequently, determined the outcome of the discussion. This finding suggests that we should try to understand whether process differs according to whether the issue was raised by the man or the woman. Again, such knowledge appears to be critical for understanding the generalizability of prediction models based on affective process.

Heyman and Hunt (this issue) suggested that our two 7-minute problem-solving tasks are rather short to have reliable estimates of negative behaviors for couples; thus, the interactions in our sample might have been less conflictual, compared to those in Gottman et al.'s (1998) Newlywed sample. This is a legitimate point and requires further investigation. Without a direct comparison, however, it would be premature to conclude that our couples showed less conflict than the Newlywed sample. Note that we conducted two different warm-up tasks (5 minutes each) before the problem-solving sessions and the discussions of men's and women's goals (7 minutes each) following the problem-solving discussions. This procedure not only allows couples to warm up before they begin to talk about problem issues in their relationships but also allows us to observe more diverse dyadic processes within dyads, including positive interaction patterns (such as support), as well as conflictual behaviors. In addition, all the men in our sample have been in our study over two decades, and many of their partners were returning participants in our Couples Study. Therefore, they were very comfortable with us and with the procedure, which made them behave more naturally during the interaction while being videotaped, compared to studies with first-time participants. As indicated by Heyman et al. (2001), the estimated time to achieve adequate reliabilities varies greatly depending on sample characteristics (i.e., distressed, nondistressed, or engaged), gender (husband vs. wife), and code (e.g., hostility, psychological abuse, withdrawal, and acceptance). The estimated time to observe reliable frequencies may also vary by task characteristics and specific coding systems. The reliability of observational data has been an issue in our field; findings based solely on observational data need to be interpreted with caution before being applied to intervention programs.

Conclusion

We argue that constructive replication attempts are seriously neglected in basic research, especially in the area of couples research, partly because such attempts are not reinforced and thus validity and reliability of many of the constructs in the literature still remain untested. We also believe that the Oregon Youth Study couples sample is representative of a much

understudied population of young couples with at-risk backgrounds and that the Kim et al. replication study provides interesting insights into affective processes in couples that will be useful for many scholars and clinicians in our field. We acknowledge that there were considerable differences between the Oregon Youth Study sample, and the Newlywed couples in Gottman et al. (1998), but do not agree that this makes our study findings, as a replication study, less valuable. We carefully presented the potential issues and limitations in our study. Further, focusing on differing sample characteristics does not generate a convincing argument as to why generalizability of findings should not have occurred. It is often the case in the social sciences that findings do not generalize, or generalize only partially from one study to another; therefore, we should be cautious of basing intervention and treatment recommendations on the basis of any single study; this is particularly the case if the study involved a relatively small sample.

Affective features of couples' interaction are an essential component in couples' relationships and strong predictors of relationship quality and stability, but whether or how affective processes within dyads vary across different socioeconomic classes or by marital status (married vs. cohabiting) has not been studied to date. Do affective processes function differently for at-risk couples? If so, how? What are the unique and common affect processes among couples with different socioeconomic backgrounds? What are the moderating contexts or factors related to dynamic processes in couples? Much work remains to be done to understand how affective processes function for couples with at-risk and low SES backgrounds, as well as couples in general. Such work would greatly inform effective intervention programs that could benefit many couples with disadvantaged backgrounds.

References

Axinn WG, Barber JS. Living arrangement and family formation attitudes in early adulthood. Journal Marriage and the Family 1997;59:595 – 611.

Axinn WG, Thornton A. The relationship between cohabitation and divorce: Selectivity or causal influence? Demography 1992;29:357 – 374. [PubMed: 1426434]

Ball FLJ, Cowan P, Cowan CP. Who's got the power? Gender differences in partner's perception of influence during marital problem solving discussions. Family Process 1995;34:303 – 321. [PubMed: 8582477]

Budetti, J.; Duchon, L.; Schoen, C.; Shikles, J. Can't afford to get sick: A reality for millions of working Americans. New York: Commonwealth Fund; 1999 Sep.

Bumpass, LL.; Sweet, JA. National Survey of Families and Households, Working Paper no. 65. Madison, WI: University of Wisconsin, Center for Demography and Ecology; 1989. Cohabitation, marriage, and nonmarital childbearing and union stability: Preliminary findings from NSFH2.

Capaldi DM, Crosby L, Stoolmiller M. Predicting the timing of first sexual intercourse for adolescent males. Child Development 1996;67:344 – 359. [PubMed: 8625717]

Carrère S, Buehlman K, Gottman JM, Coan J, Ruckstuhl L. Predicting marital stability and divorce in newlywed couples. Journal of Family Psychology 2000;14:42 – 58. [PubMed: 10740681]

Clarkberg M. The price of partnering: The role of economic well-being in young adults' first union experiences. Social Forces 1999;77:945 – 968.

Coan JA, Gottman JM. Sampling, experimental control, and generalizability in the study of marital process models. Journal of Marriage and Family. this issue

Cohan CL, Kleinbaum S. Toward a greater understanding of the cohabitation effect: Premarital cohabitation and marital communication. Journal of Marriage and Family 2002;64:180 – 192.

Cooper M. District attorney expects crime rate rise. The Register-Guard. 2006 September 19;

Dush CMK, Cohan CL, Amato PR. The relationship between cohabitation and marital quality and stability: Change across cohorts? Journal of Marriage and Family 2003;65:539 – 549.

Elliott, DS.; Huizinga, D.; Morse, BJ. The dynamics of delinquent behavior: A National Survey Progress Report. Boulder, CO: Institute of Behavioral Sciences, University of Colorado; 1985.

Fagan, J.; Browne, A. Violence between spouses and intimates: Physical aggression between women and men in intimate relationships. In: Reiss, AJ.; Roth, JA., editors. Understanding and preventing violence: Social influences. 3. Washington, DC: National Academy Press; 1994. p. 115-292.

Gottman JM, Coan J, Carrère S, Swanson C. Predicting marital happiness and stability from newlywed interactions. Journal of Marriage and the Family 1998;60:5 – 22.

Hawkins MW, Carrère S, Gottman JM. Marital sentiment override: Does it influence couples' perceptions? Journal of Marriage and Family 2002;64:193 – 201.

Heyman RE. Observation of couple conflicts: Clinical assessment applications, stubborn truths, and shaky foundations. Psychological Assessment 2001;13:5 – 35. [PubMed: 11281039]

Heyman RE, Chaudhry BR, Treboux D, Crowell J, Lord C, Vivian D, Waters EB. How much observational data is enough? An empirical test using marital Interaction coding. Behavior Therapy 2001;32:107 – 123. [PubMed: 16733528]

Heyman RE, Hunt AN. Replication in observational couples research: A commentary. Journal of Marriage and Family. this issue

Kline GH, Stanley SM, Markman HJ, Olmos-Gallo PA, Peters M, Whitton SW, Prado LM. Timing is everything: Pre-engagement cohabitation and increased risk for poor marital outcomes. Journal of Family Psychology 2004;18:311 – 318. [PubMed: 15222838]

Manning WD, Smock PJ. First comes cohabitation and then comes marriage? A research note. Journal of Family Issues 2002;23:1065 – 1087.

Mihalic, S.; Irwin, K.; Elliott, DS.; Fagan, A.; Hansen, D. OJJDP Juvenile Justice Bulletin. Washington, DC: U.S. Department of Justice, Office Justice Programs, Office of Juvenile and Justice Delinquency Prevention.; 2001. Blueprints for violence prevention.

Moffitt, TE.; Caspi, A. Findings about partner violence from the Dunedin Multidisciplinary Health and Development Study. Washington, DC: U.S. Department of Justice, Office of Justice Programs, National Institute of Justice; 1999. (NCJ 170018)

Navratil JL, Green SM, Loeber R, Lahey BB. Minimizing subject loss in a longitudinal study of deviant behavior. Journal of Child and Family Studies 1994;3:89 – 106.

Seltzer, JA. Families formed outside of marriage. In: Milardo, RM., editor. Understanding families into the new millennium: A decade in review. Minneapolis, MN: National Council on Family Relations; 2001. p. 466-487.

Smock PJ. Cohabitation in the United States: An appraisal of research themes, findings, and implications. Annual Review of Sociology 2000;26:1 – 20.

Stanley SM, Whitton SW, Markman HJ. Maybe I do: Interpersonal commitment and premarital or nonmarital cohabitation. Journal of Family Issues 2004;25:469 – 519.

Thomson E, Colella U. Cohabitation and marital stability: Quality or commitment? Journal of Marriage and the Family 1992;54:259 – 267.

Table 1

Means Levels of Affect Constructs by Relationship Status

	Married (*n* = 28)	Cohabiting (*n* = 57)	*p*	Stably married (*n* = 22)	Stably cohabiting (*n* = 32)	*p*
Men's anger	7.21	8.98	.64	6.50	11.38	.36
Women's anger	11.36	16.47	.39	13.18	14.84	.80
Men's low-intensity negative affect	122.54	94.63	.27	130.36	110.03	.57
Women's low-intensity negative affect	169.68	143.12	.32	170.23	135.44	.30
Men's high-intensity negative affect	51.93	50.42	.87	50.27	53.38	.76
Women's high-intensity negative affect	56.29	47.51	.36	52.27	47.66	.66
Men's positive affect	71.89	71.54	.98	72.18	72.53	.98
Women's positive affect	76.32	67.82	.51	79.32	69.28	.54
Men's positive/positive + negative affect	0.37	0.37	.92	.36	.38	.76
Women's positive/positive + negative affect	0.31	0.32	.94	.31	.35	.65
Men's reciprocity of women's low-intensity negative affect	28.79	21.82	.42	28.64	25.13	.77
Women's reciprocity of men's low-intensity negative affect	28.64	22.26	.45	28.68	25.25	.78
Men's reciprocity of women's high-intensity negative affect	5.68	4.37	.45	4.45	4.00	.76
Women's reciprocity of men's high-intensity negative affect	4.50	4.42	.96	3.14	4.25	.44
Men's negative escalation	14.43	11.07	.31	12.95	9.66	.26
Women's negative escalation	11.43	7.18	.16	11.41	8.13	.44
Men's negative start up	101.11	91.63	.60	110.32	106.97	.89
Women's negative start up	150.96	130.09	.34	152.91	121.06	.25
Men's deescalation of women's low-intensity negative affect	114.46	98.37	.39	117.91	89.56	.23
Women's deescalation of men's low-intensity negative affect [Indent second line of row header 3 spaces.]	72.21	58.42	.39	79.23	69.28	.65
Men's deescalation of women's high-intensity negative affect [Indent second line of row header 3 spaces.]	35.04	32.18	.62	32.91	31.72	.86
Women's deescalation of men's high-intensity negative affect [Indent second line of row header 3 spaces.]	27.75	30.32	.66	30.59	34.56	.60