Personality and Social Psychology Review
2002, Vol. 6, No. 4, 326–340

A General Systems Theory of Marriage: Nonlinear Difference Equation Modeling of Marital Interaction

John Gottman, Catherine Swanson, & Kristin Swanson
Department of Psychology
University of Washington

This article describes a new mathematical approach for modeling the prediction of divorce or marital stability from marital interaction using nonlinear difference equations. The approach is quite general for modeling social interaction, and can be applied to any time series data generated over time for two individuals. We pursued a balance model in selecting the dependent variables of this modeling. Both the mathematical methods and the theoretical gains obtained when using this approach are reviewed.

What is the Modeling Trying to Accomplish?

In our laboratory's multi-method research on marital interaction we have shown, in four longitudinal studies, that we can predict with over 90% accuracy whether a couple will divorce or stay married, and their marital satisfaction if they do stay married (Gottman, 1994; 1999). Furthermore, we were able to predict in three measurement domains: the couple's interactive behavior, their perception of their interaction, and their physiology during the interaction. The prediction in the behavioral domain was based largely on coding positive and negative emotions.

What we started doing nine years ago, in collaboration with the mathematical biologist James Murray and his students, was to build a mathematical model for these predictions. Unlike other problems in mathematical biology, in which it was possible to write the differential equations from existing theory, we had no laws and no mathematical theory in the area of marriage. For that reason the goal of our efforts became the *construction of theory.* We wound up developing a new language for social interaction and a theory that utilizes that language to attempt to understand our predictions.

The mathematically-based theory we developed made it possible to simulate the couple's behavior under conditions we had never observed them in, and so it led to the idea of conducting proximal change experiments. Instead of intervening to change the entire marriage, these proximal change experiments had as their goal changing a particular parameter of the mathematical model for that couple. Thus, the study of marriage was brought into the social psychology laboratory. This article is an introduction to the work we did to build that mathematical theory. First, let us briefly review research on marriage and marital interaction.

Requests for reprints should be sent to John Gottman, James Mifflin Professor, University of Washington, Box 351525, Department of Psychology, Seattle, WA 98195. E-mail: jgottman@gottman.com

Observational Research on Marital Relationships: Brief Review

Psychology was a late comer to the study of marriage. Sociologists had been studying marriages for 35 years before psychologists got wind of the idea, although the first published study on marriage was conducted by a psychologist, Louis Terman in 1938 (Terman, Buttenweiser, Ferguson, Johnson, & Wilson, 1938). Terman's question was "What are the correlates of marital happiness and misery?" What psychologists initially brought to the study of marriage in the 1970s was the use of observational methods, the design and evaluation of intervention programs, and an unbridled optimism. These contributions have had an enormous impact on the study of marriage. Psychology was initially skeptical about studying marriage, in part because personality theory was facing a severe challenge in the 1970s from the work by Mischel (1968). In that book, Mischel reviewed research on personality and suggested that personality theory had come far short of being able to predict and understand behavior. He concluded that correlations were quite low on the whole, that the field was plagued with common method variance (mostly self-reports predicting self-reports), and that the best predictors of future behavior were past behavior in similar situations. This book was a great stimulus to many researchers. It encouraged a new look at personality measurement, validity, and reliability (Wiggins, 1973). It stimulated new kinds of research in personality. However, it also contributed to a pessimistic view that research in interpersonal psychology would have very little payoff. In hindsight, this pessimism was wrong. We are learning that, in fact, much of the order in individual personality exists at the interpersonal level. Patterson (1982), in his conclusion that there is a great deal of consistency across time and situations in aggression, suggested that the aggressive trait ought to be rethought in interpersonal terms as the aggressive boy's recasting people in his social world to play out dramatic coercive scenes shaped in his family. The same is true of gender differences: they appear to emerge primarily in the context of relationships (Maccoby, 1990).

The use of self-report measures, including personality measures, had initially dominated the field of marriage research. Unfortunately, even with the problem of common method variance, the self-report paper-and-pencil personality measures yielded relatively weak correlates of marital satisfaction (Burgess, Locke, & Thomes, 1971). Not until studies asked spouses to fill out questionnaires about their spouse's personality, were substantial correlations discovered with marital satisfaction. Unhappily married couples were found to endorse nearly every negative trait as characteristic of their spouses (the negative halo effect), while happily married spouses were found to endorse nearly every positive trait as characteristic of their spouses (the positive halo effect; Nye, 1988).

A complete review of findings on marriage is beyond the scope of this article. We will limit ourselves to a sampler, and some general conclusions. Because we are interested in temporal patterns of behavior, perception, and physiology that unfold over time, we will restrict ourselves to observational studies that have included some sequential analyses of the data. This restricts us to the work of seven laboratories, Weiss's (Oregon), Raush's (Massachusetts), Gottman's (Washington), Schapp's (Holland), Ting-Toomey's (New Jersey), the Max Planck group in Munich (Revenstorf, Hahlweg, Schindler, Vogel), and Fitzpatrick's (Wisconsin). What were the results of various laboratories that investigated the Terman question, and in particular what were the results of sequential analyses? Of the many studies that have observed marital interaction, very few have employed sequential analyses. With the Marital Interaction Coding System (MICS), there are only two such studies (Margolin & Wampold, 1981; Revenstorf, Vogel, Wegener, Hahlweg, & Schindler, 1980). They collapsed the many codes of the MICS into positive or negative, or into positive, negative, and neutral. They each defined negativity in their own way. Margolin and Wampold (1981) reported the results of interaction with 39 couples (combined from two studies conducted in Eugene, Oregon and Santa Barbara, California). Codes were collapsed into 3 global categories: Positive (problem-solving, verbal and nonverbal positive), Negative (verbal and nonverbal negative), and Neutral.

What were the results? Margolin and Wampold's (1981) results on negative reciprocity were that distressed couples showed negative reciprocity through Lag 2, whereas nondistressed couples do not demonstrate it to any significant extent. For positive reciprocity, they found that "whereas both groups evidenced positive reciprocity through Lag 2, this pattern appears to continue even into Lag 3 for distressed couples" (p. 559). Thus, reciprocating positive acts was more likely for distressed than for nondistressed couples. Gottman (1979) had reported similar results, suggesting that distressed couples showed greater rigidity in temporal interactional structure than nondistressed couples. Margolin and Wampold also defined a sequence called "negative reactivity," which involves a positive response to a negative antecedent by one's spouse. They proposed that there is a suppression of positivity following a negative antecedent in distressed couples. They found this for all four lags for distressed coupes, but they found no evidence for this suppression of positivity by negativity for any lag for nondistressed couples.

Revenstorf, et al. (1980), studying 20 German couples, collapsed the MICS categories into six rather than three summary codes. These codes were Positive Reaction, Negative Reaction, Problem Solution, Problem Description, Neutral Reaction, and Filler. Interrupts, Disagrees, Negative Solution, and Commands were considered Negative. They employed both lag sequential analyses that allowed them to examine sequences out for four lags, as well as the multivariate information theory that Raush, Barry, Hertel and Swain (1974) had employed with couples undergoing the transition to parenthood. From the multivariate information analysis, they concluded

> In problem discussions distressed couples respond differently from non-distressed couples....In particular [distressed couples] are more negative and less positive following positive (+) and negative (-) reactions. At the same time they are more negative and more positive, that is more emotional, following problem descriptions (P) of the spouse. Above all distressed couples are more negative and less positive in general that non-distressed couples. (p.103)

They also found 17 sequences that differentiated the two groups. There is some inconsistency in the group differences for sequences with similar names (e.g., "reconciliation"), here only their clearest results will be summarized. For what might be called "constructive" interaction sequences, they found that nondistressed couples engaged in more "validation" sequences (problem description followed by positivity), and positive reciprocity sequences (positive followed by positive). On the "destructive side," they found that distressed couples engaged in more "devaluation" sequences (negative follows positive), negative continuance sequences (which they called "fighting on" or "fighting back" in 3-chain sequences) and negative startup sequences (which they called "yes-butting," meaning that somewhere in the 4-chain sequence, negative follows positive) than nondistressed couples. After an analysis of the sequences following a problem description, they concluded

> It appears as if the distressed couples would interact like non-distressed—had they only higher positive response rates following a problem description of the spouse. And vice versa. The non-distressed would react equally detrimentally as the distressed—were they to respond more negatively to problem description of their spouse. The way they handle problems [problem description statements] seems to be the critical issue —not the sheer number of problems stated. (p.107)

Revenstorf et al. (1980) also continued their sequential analyses for five lags and found that these reciprocity differences held across lags. They wrote:

> In summary, different patterns of response tendencies emerge for distressed and non-distressed couples. After a positive statement the partner continues to reciprocate it positively in non-distressed, whereas no immediate response is likely in distressed couples. After a negative statement no immediate response is most likely in non-distressed, whereas in distressed couples both partners continue to reciprocate negatively. A problem description finally is repeatedly followed by a positive response in non-distressed. In distressed couples, negative statements follow repeatedly. (p. 109)

Revenstorf et al. then described four types of sequences. The first type of sequence is continued negativity (they called it "distancing"). This sequence measures the extent to which negativity becomes an absorbing state. The second sequence type was positive reciprocity (which they called "attraction"). This sequence measures the extent to which positivity becomes an absorbing state. The third sequence consisted of alternating problem descriptions and negativity (they called it "problem escalation"). The fourth type of sequence consists of Validation sequences, sequences of alternating problem descriptions and positive responses to it (they called it "problem acceptance"). In most of their graphs (e.g., for positive reciprocity), the differences between the groups were not very great. However, the evidence was very clear that negativity represented an "absorbing state" for distressed couples, but not for non-distressed couples. By Lag-2, non-distressed couples begin to escape from the negativity, but distressed couples can not escape. These graphs provide dramatic information of group differences reflected in sequential patterning of MICS codes.

The consistent findings in these two and other studies that have employed sequential analysis (Fitzpatrick, 1988; Gottman, 1979; Rausch et al., 1974; Schaap, 1982; Ting-Toomey, 1982; for a review, see Gottman, 1994) is that: (1) unhappily married couples appear to engage in long chains of reciprocated negativity, and (2) there is a climate of agreement created in the interaction of happily married couples.

These findings using observational research on marital interaction are interesting and important. However, they are also frustrating because of the absence of theory either in generating or in summarizing the empirical findings. Just where is all this empirical research heading? What is it saying about dysfunctional and functional marriages?

Another dust-bowl empirical approach was taken by Gottman and Levenson (1992). They used a graphical method for combining categorical observational data to produce husband and wife *time-series data* over a 15-minute conflict interaction. They developed a methodology for obtaining synchronized physiological, behavioral, and self-report data in a sample of 73 couples who were followed longitudinally, initially between 1983 and 1987. Using observational coding of interactive behavior with the Rapid Couples Interaction Scoring System (Gottman, 1996), couples were divided into two groups, which we here call "high risk," and "low risk" for divorce. This classification was based on a graphical method originally proposed by Gottman (1979) for use with the Couples Interaction Scoring System, a predecessor of the RCISS, for coding how couples attempt to resolve a marital conflict. This observational system used as its unit the "interact," which is everything each of two people say in two conversational turns at speech during a 15-minute task in which they attempt to resolve a major area of continuing disagreement in their marriage. This unit provided an uneven number of interacts for each couple. The coding also took about 6 hr, and an additional 10 hr to create a necessary verbatim transcript. We have now been able to derive essentially the same data using a weighting of the codes of the Specific Affect Coding System (SPAFF; Gottman, 1996), which can be coded on-line almost in real time, and, with a numerical weighting scheme for each coding category, averaged over –second blocks, which provides 150 observations for each person in each couple. SPAFF also has the advantage that it can be coded for any conversation, not just for conflict resolution conversations.

In the initial data, on each conversational turn the total number of positive RCISS speaker codes minus the total number of negative speaker codes was computed for each spouse. The classifications of "positive" or "negative" were based on previous studies that had sought to discriminate happy from unhappy couples from a scoring of the behavior they exhibited during conflictual marital interaction. Based on this review of the literature, the RCISS system was devised. Then the cumulative total of these points was plotted for each spouse. The slopes of these plots, which were thought to provide a stable estimate of the difference between positive and negative codes over time, were determined using linear regression analysis. If both husband and wife graphs had a positive slope, they were called "low risk;" if not, they were called "high risk." This classification is referred to as the couple's "affective balance." All couples, even happily married ones, have some amount of negative interaction; similarly, all couples, even unhappily married, have some degree of positive interaction.

There was some implicit theory in this dust-bowl empirical approach. Computing the graph's slope was guided by a *balance theory* of marriage, namely that those processes most important in predicting dissolution would involve a balance, or a *regulation*, of positive and negative interaction. Low risk couples were defined as those for whom both husband and wife speaker slopes were significantly positive; high risk couples had at least one of the speaker slopes that was not significantly positive. By definition, low risk couples were those who showed, more or less consistently, that they displayed more positive than negative RCISS codes. Classifying couples in the current sample in this manner produced two groups consisting of 42 low risk couples and 31 high risk couples. We can easily imagine two cumulative graphs, one from the interaction of a low risk and one the interaction of a high risk couple.

In 1987, four years after the initial assessment, the original subjects were recontacted and at least one spouse (70 husbands, 72 wives) from 73 of the original 79 couples (92.4%) agreed to participate in the follow-up. Marital status information was obtained. Gottman and Levenson then used their longitudinal data to examine the results for the dissolution variables of their "dissolution cascade." The dissolution cascade is a Guttman scale in which precursors of separation and divorce were identified as continued marital unhappiness and serious thoughts of dissolution. Based only on the Time-1 graphs, they found that after four years low risk couples were indeed less likely to be unhappy, to have persistent thoughts of divorce, less likely to be lonely in the marriage, less likely to lead "parallel lives" (avoiding one another), and less likely to separate and divorce than high risk couples. We have now followed the original sample for 14 years.

Let us take a step back from these empirical findings and think about how dynamical systems theory might be employed to head our research in a more theoretical direction.

The General Systems Theory of von Bertalanffy

The application of applied mathematics to the study of marriage was presaged by von Bertalanffy (1968), who wrote a classic and highly influential book called *General System Theory.* This book was an attempt to view biological and other complex organizational units across a wide variety of sciences in terms of the interaction of these units. The work was an attempt to provide a holistic approach to complex systems. The work fit a general Zeitgeist. As von Bertalanffy noted, his work fit with Wiener's (1948) *Cybernetics*, Shannon & Weaver's information theory, and VonNeumann and Morgenstern's (1947) game theory. The concepts of homeostasis (derived from the physiologist Walter Cannon), feedback, and information provided the basis for a new approach to the study of complex interacting systems.

The mathematics of general system theory was lost to most of the people in the social sciences who were inspired by von Bertalanffy's work. Our work is, therefore, a return to von Bertalanffy's original dream. Von Bertalanffy's dream was that the interaction of complex systems with many units could be characterized by a set of values that changed over time, denoted Q_1, Q_2, Q_3, and so on. We can presume that each Q variable indexed a particular unit in the "system," such as mother, father, and child, and, furthermore that these variables measured some relevant characteristic of a person that changes over time, such as the number of angry facial expressions per unit time. Actually, the Q's were quantitative variables that von Bertalanffy never specified.

However, he thought that the system could be best described by a set of ordinary differential equations of the form:

$$dQ_1/dt = f_1(Q_1, Q_2, Q_3, \ldots)$$
$$dQ_2/dt = f_2(Q_1, Q_2, Q_3, \ldots)$$
and so on.

The terms on the left of the equal sign are time derivatives, that is, rates of change of the quantitative sets of values Q_1, Q_2, Q_3, and so on. The terms on the right of the equal sign are functions, $f_1, f_2, \ldots$, of the Q's. Von Bertalanffy thought that these functions, the f's, would generally be nonlinear. The equations he selected have a particular form, called "autonomous," meaning that the f's have no explicit function of time in them, except through the Q's, which are functions of time. These are, in fact, the type of equations we worked with. However, von Bertalanffy presented a table in which these nonlinear equations were classified as "impossible" (von Bertalanffy, 1968, p.20). He was referring to a very popular mathematical method of approximating nonlinear functions with a linear approximation.

However, it was not the case that these systems were "impossible"; von Bertalanffy was unaware of the extensive mathematical work beginning in the 19th century with Poincaré, on nonlinear differential equations, chaos, and fractal theory, which was only to become known to the popular press in the 1980s. In fact, in recent times the modeling of complex deterministic (and stochastic) systems with a set of nonlinear difference or differential equations has become a very productive enterprise across a wide set of phenomena, across a wide range of sciences, including the biological sciences.

Nonlinear Dynamic Modeling With More Than One Equation

We thus applied a relatively old approach to the new problem of modeling social interaction using the mathematics of difference and differential equations. These equations express, in mathematical form, a proposed mechanism of change over time. They do not represent a statistical approach to modeling but rather they are designed to suggest a precise mechanism of change. This method has been employed with great success in the physical and the biological sciences (e.g., see Murray, 1989). It is a quantitative approach that requires the modeler to be able to write down, in mathematical form and on the basis of some theory, the causes of change in the dependent variables. For example, in the classic predator-prey problem, one writes down that the rate of change in the population densities is some function of the current densities. The equations are designed to write down the precise form of rates of change over time. The ideal mathematical technique for describing change is the area of differential equations. These equations usually used linear terms or linear approximations of nonlinear terms, and they often gave very good results. In fact, most of the statistics psychology uses are based upon linear models. In the area of differential equations, linear equations simply assume that rates of change follow generalized straight-line functions of the variables rather than curved line functions.[1] Unfortunately, linear models are generally unstable.

In recent years it has become clear that most systems are complex and must be described by nonlinear terms. Interestingly, by employing nonlinear terms in the equations of change, some very complex processes can be represented with very few parameters. Unfortunately, unlike many linear equations, these nonlinear equations are generally not solvable in closed functional mathematical form. For this reason the methods have been called "qualitative," and visual methods must be relied upon. For this purpose, numerical and graphical methods have been developed such as "phase space plots." These visual approaches to mathematical modeling can be very appealing because they can engage the intuition of a scientist working in a field that has no mathematically stated theory. If the scientist has an intuitive familiarity with the data of the field, our approach may suggest a way of building theory using mathematics in an initially qualitative manner. The use of these graphical solutions to nonlinear differential equations makes it possible to talk about "qualitative" mathematical modeling. In qualitative mathematical modeling, one searches for solutions that have similarly shaped phase space plots. We will now describe these methods of mathematical modeling in detail.

The First Question: What are the Steady States of the System?

Once we write down the equations of marital interaction, the first question is "Toward what values is the system drawn?" To answer the question we define a "steady state" as one for which the derivatives (on the left side of the von Bertalanffy equations) are zero. This means that the system at a steady state does not change.

The Second Question: Which Steady States are Stable?

What does "stability" mean? It means that if you perturb the system slightly from a stable steady state, it will return to that steady state. It is as if the steady state is an *attractor* that pulls the system back to the steady state. This is like a rubber band snapping back once pulled and released. If the steady state is unstable, on the other hand, and you perturb the system slightly at that steady state, it will move away from that steady state. If we have the more general equation $N' = dN/dt = f(N)$, it is easy to show that N^* is a stable steady state if $f' = df/dt$ is less than zero at N^* and unstable if $f' = df/dt$ is positive at N^*. Graphically we can look at the slope of f(N) where it crosses the x-axis; each point of negative slope would imply a stable steady state, while each point of positive slope would imply an unstable steady state. Recall that the function f(N) is given by our equation, once we know how to write it down. In our work with steady states, we will refer to the stable steady states as "set points," a term borrowed from research on the management of body weight. In this area, it has been noted that the body defends a particular weight as a set point, adjusting metabolism to maintain the body's weight in homeostatic fashion. It is interesting that the existence of homeostasis does not, by itself, imply that the system is regulated in a functional manner. Recently a man who weighed 1,000 pounds died in his early thirties; he had to be lifted out of his apartment with a crane. Clearly this set point was dysfunctional for this man's health, despite the fact that his body defended this set point.

Describing the Behavior of the Model

The next step in modeling is to describe the behavior of the model near each steady state and as the parameters of the model vary. We want to know what the model tells us qualitatively about how the system is supposed to act. Then, if the model isn't acting the way we think it ought to (given the phenomena we are try-

[1]The strength of an attractor is one over the smallest eigenvalue for that state. Eigenvalues are solutions to two equations when linearizing around a steady state.

ing to model), we alter the model by changing the function "f" . This process can be represented with a flow chart for dynamic modeling of any process over time. We begin with identifying the phenomenon or phenomena we wish to model. This is given to us by our insights and intuitions about the science. Then we write the equations for building the model. This is a difficult task, one that requires some knowledge of mathematics. Next, we find the steady states of the model, those points where the derivatives are zero, and determine if these steady states are stable or unstable. Then we study the qualitative behavior of the model near the steady states. Next, we study how the model behaves as we vary the parameters of the model. Finally, returning to the science, we ask whether this model is doing the job we want, and, if not, we modify the model. There is sometimes an additional step in the model building, namely, non-dimensionalizing the model. This step reduces the number of parameters in the model.

Writing the Equations of Marital Interaction

In modeling marital interaction there are two equations, one for the husband, and one for the wife. Our dependent variable was the positive minus negative at each turn of speech. Because we could not come up with any theory we knew of to write down the equations of change (linear or nonlinear) in marital interaction over time, we developed an approach that uses both the data and the mathematics of differential or difference equations in conjunction with the creation of qualitative mathematical representations of the forms of change. The expressions we wrote down were then used with the data to "test" our qualitative forms. What we discovered was different about this approach was that we needed to use the modeling approaches to generate the equations themselves. Thus, the objectives of the mathematical modeling in our case became to generate theory. We think that our experience is quite general and may also be useful for other social psychological problems.

The Details of our Dynamic Mathematical Modeling

The goal of our modeling was to dismantle the unaccumulated RCISS point graphs of positive minus negative behaviors at each turn into components that had some theoretical meaning. This is an attempt at understanding the ability of these data to predict marital dissolution via the interactional dynamics. We thus began with the Gottman-Levenson dependent variable for husband and for wife and dismantled it into components that represent: (1) a function of interpersonal influence from spouse to spouse, and (2) terms containing parameters related to an individual's own dynamics. This dismantling of the Gottman-Levenson dependent variable into "influenced" and "uninfluenced" behavior represents part of our theory of how the dependent variable may be decomposed into components that suggest a mechanism for the successful prediction of marital stability or dissolution.

The qualitative and theoretical portion of writing our equations lies in writing down the mathematical form of the influence functions. An influence function is used to describe the couple's interaction. The mathematical form is represented graphically, with the x-axis as the range of values of the dependent variable (positive minus negative at a turn of speech) for one spouse and the y-axis the average value of the dependent variable for the other spouse's immediately following behavior, averaged across turns at speech. This latter point is critical, and it may be unfamiliar to social scientists: The influence functions represent averages across the whole interaction.

Now what did we know about marital interaction that could help us write down the mathematical form of the influence functions? Gottman (1994) suggested that one consistent result that had been obtained by many laboratories studying marital conflict interaction with observational methods was that negative affect was a better correlate of marital satisfaction and predictor of longitudinal course than positive affect. This means that we could expect that the theoretical form of the influence functions would probably be bilinear, with a steeper slope in the negative affect ranges than in the positive affect ranges (this is also reminiscent of Alexander's defensive/supportive cycle in family interaction; e.g., Alexander, 1973). Thus, we would expect the influence functions to be somewhat asymmetric, as shown in Figure 1.

The Parameters

Notice in Figure 1 that the *slopes of the influence functions* in the two regions are the important parameters for the bilinear form of the influence functions. If

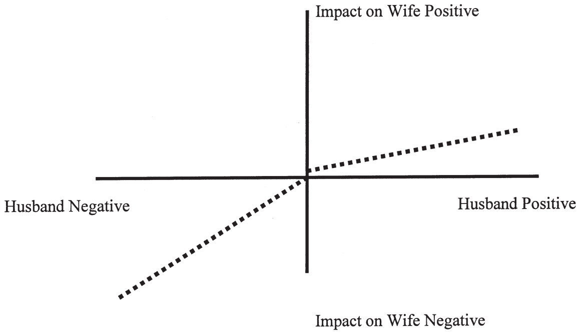

Figure 1. Bilinear form of the influence functions. This form is based on the frequent result that negative affect during conflict is more strongly related to marital satisfaction than positive affect.

we had selected as our theoretical influence functions an ojive form (step functions instead of straight lines) we could have had two parameters indicating the heights of each part of the O-Jive; furthermore, we would have also added two parameters, the thresholds of positivity and negativity. These thresholds represent the values of how negative the husband's negativity has to get before it starts having an impact on the wife, and how positive the husband's behavior has to get before it starts having a positive impact on the wife. But in the bilinear model only two parameters are obtained for each spouse's influence on their partner. The choice of the influence function thus determines the nature of the theory we will build. We have experimented with the ojive form of the influence function (see Gottman, Murray, Swanson, Tyson, & Swanson, in press).

Using the existing research on marital interaction, another parameter we decided that it was important to include is the emotional inertia (positive or negative) of each spouse, which is each person's tendency to remain in the same state for a period of time. The greater the emotional inertia, the more likely the person is to stay in the same state for a longer period of time. It has been consistently found, in marital interaction research, that the reciprocation of negativity is more characteristic of unhappy than of happy couples. This finding has held cross-nationally as well as within the United States (for a review, see Gottman, 1994). Surprisingly, the tendency to reciprocate positive affect is also greater in unhappy than in happy couples (see Gottman, 1979). There is generally more time linkage or temporal structure in the interaction of distressed marriages and families. Another way of stating this finding is that there is more new information in every behavior in well-functioning family systems. The system is also more flexible because it is less time-locked. A high inertia spouse is also less open to being influenced by the partner. Emotional inertia came from including the autocorrelation component of human behavior.

Another parameter we added after four years of working on the model was a constant that represented the initial starting values of the conversation. A derived parameter from knowing both this starting value and the emotional inertias of both people (one that emerged from solving the equations) was the couple's uninfluenced set point, which is their average level of positive minus negative when their spouse did not influence them (the influence function was zero). We decided that this state of affairs was most likely when the affect was most neutral or equally positive and negative. We think of this parameter as what each spouse brings into the interaction, before being influenced by the partner. Clearly this uninfluenced set point may be a function of the past history of the couple's interactions. It may also be a function of individual characteristics of each spouse, such as a tendency to dysphoria, or optimism.

Another derived parameter was the influenced set point of the interaction, which is a steady state of the system. One way of thinking about the influenced set point is that it is a sequence of two scores (one for each partner) that would be repeated ad infinitum if the theoretical model exactly described the time series; if such a steady state is stable, then sequences of scores will approach the point over time. In a loose sense it represents the average score the theoretical model would predict for each partner. We also thought it might be interesting to examine the difference between influenced and uninfluenced set points. We expect that the influenced set point will be more positive than the uninfluenced set point in marriages that are stable and happy; that is, we asked the question, Did the marital interaction pull the individual in a more positive or a more negative direction? This was an additional derived parameter in our modeling.

We think it will be interesting to discover whether some interventions will alter either the fundamental shape of the influence function and whether other interventions will alter the influenced or uninfluenced set points. Changing the influence functions seems to us to represent much more fundamental change of the relationship itself than changing the uninfluenced set points.

Estimating the Parameters

We needed a plan for estimating these and other salient parameters. We begin with a sequence of Gottman-Levenson scores: W_t, H_t, W_{t+1}, H_{t+1},… For purposes of estimation we assumed that zero scores have no influence on the partner's subsequent score. We then subtract the uninfluenced effects from the entire time series to reveal the influence function, which summarizes the partner's influence. We then can plot the influence function as a function of values of the data (positive minus negative affect), and use the theoretical form of the influence function to estimate the two slopes. There are two influence functions, the influence of the husband on the wife and the influence of the wife on the husband. We can also estimate the couple's relative power by subtracting the two sets of slopes from one another, comparing the wife's and the husband's influence. Notice we have complicated the usual discussions of relative power, because relative power here depends on the affect.

Description of the Model

The simplifying assumption that each person's score is determined solely by their own and their partner's previous score restricts us to a particular class of mathematical models. It we denote W_t and H_t as the husband's and wife's scores respectively at turn t, then the sequence of scores is given by an alternating pair of coupled difference equations:

$$W_{t+1} = f(W_t, H_t)$$
$$H_{t+1} = g(W_{t+1}, H_t) \qquad (1)$$

The functions *f* and *g* remain to be determined. The asymmetry in the indexes is due to the fact that we are assuming, without loss of generality, that the wife speaks first. We therefore label the turns of speech W_1, H_1, W_2, H_2,… To select a reasonable *f* and *g*, we make some simplifying assumptions. First we assume that the past two scores contribute separately and that the effects can be added together. Hence, a person's score is regarded as the sum of two components, one of which depends on their previous score only and the other on the score for their partner's last turn of speech. We term these the "uninfluenced" and the "influenced" components, respectively.

Consider the uninfluenced component of behavior first. This is the behavior one would exhibit if not influenced by one's partner. As we noted, it could primarily be a function of the individual, rather than the couple, or it could be a cumulative effect of previous interactions, or both. It seems reasonable to assume that some people would tend to be more negative when left to themselves while others would naturally be more positive in the same situation. This "baseline temperament" we term the individual's "uninfluenced set point." We suppose that each individual would eventually approach that set point after some time regardless of how happy or how sad they were made by a previous interaction. The simplest way to model the sequence of uninfluenced scores is to assume that uninfluenced behavior can be modeled by a simple linear relationship. This leads us to the linear relationship:

$$P_{t+1} = r_i P_t + a_i \qquad (2)$$

where P_t is the score at turn *t*, r_i determines the rate at which the individual returns to the uninfluenced set point and a_i is a constant. The constant r_i will henceforth be referred to as the "emotional inertia," parameter, or more simply, just the "inertia."

The uninfluenced set point is the steady state of this equation and is found by solving $P_{t+1} = P_t = P = a_i / (1\text{-}r_i)$. This is the *uninfluenced steady state* , that is, the attractor of the system if there were no influence. Note that the stability of the attractor in Equation 2 is governed by the value of r_i. If the absolute value of r_i is less than 1.0, then the system will tend toward the steady state regardless of the initial conditions, while if the absolute value of r_i is greater than 1.0, the system will always evolve away from a steady state. Clearly we require the uninfluenced steady state to be stable, and so we are only interested in the case in which the absolute value of r_i is less than 1.0. The magnitude of r_i determines how quickly the uninfluenced state is reached from some other state, or how easily a person changes their frame of mind, hence the use of the word "inertia." For selecting the form of the influenced component of behavior, we can take several approaches. The influence function is a plot of one person's behavior at turn, t, on the x-axis, and the subsequent turn, t+1, behavior of the spouse on the y-axis. Averages are plotted across the whole interaction. Our approach was to write down a theoretical form for these influence functions (recall Figure 1). As we noted, we posited a two-slope function: We have two straight lines going through the origin, with two different slopes, one for the positive range and one for the negative range. Other forms of the influence function are also reasonable.

The Full Equations

We denote the influence functions by I_{AB} (A_t), the influence of person A's state at turn t on person B's state. With these assumptions the complete model is:

$$W_{t+1} = I_{HW}(H_t) + r_1 W_t + a \qquad (3)$$

$$H_{t+1} = I_{WH}(W_{t+1}) + r_2 H_t + b \qquad (4)$$

Again, the asymmetry in the indexes is due to the fact that we are assuming that the wife speaks first. The problem now facing us is estimation of our four parameters, r_1, a , r_2, and b, and the empirical determination of the two unknown influence functions.

Proximal Change Experiments: Changing Marriages Through Model Simulation and Intervention

It is very important that the model be derived in such a manner that the parameters and functions of the model have interpretable physical meaning. For example, the model parameters *a* and *b* can be interpreted as startup values before autocorrelation (self-influence) or partner influence begins. Once we have a model, with equations, and we have estimates of the model's parameters, we can simulate the couple's interaction under conditions different from those that were used in the estimation. This means that we can change the model's parameters in ways that are meaningful in the sense that they can be translated into behavior. For example, suppose we have a couple whose conflict interaction we model, and we find that the husband's startup parameter is negative, and, furthermore, we find that the model has only a negative attractor (stable steady state). Now let us simulate the model's behavior if the husband's startup parameter, *b*, were far more positive. This is done simply by beginning with the same equations (with only one parameter changed) and the same initial values, and running off the predicted values. We calculate the new influence functions, and we once again compute the new stable steady states. Suppose we find that now the model has a positive stable steady state. That means that the model has predicted that changing the husband's start values will have a major positive effect on the relationship. What we now need to do is to create an intervention that has the desired effect on the husband. Then the couple actually has the post-intervention conversation, and the model is refit. The question then is, does the new model look like the simulated model? In that way we can use the modeling and the simulation to test the model. For the past six years we have pilot tested these proximal change experiments and they have greatly enriched our understanding of how to change marriages. The mathematical modeling has been an essential ingredient in these experiments.

Estimation of Parameters and the Unknown Influence Functions

The algorithms and computer program for using our modeling will soon be available upon request. We will now describe the general methods we have developed for estimation of model parameters and the influence functions. We begin by examining the model for that subset of data points for which we can safely assume that there was no partner influence. We assumed that these points were those for which the partner's score was near zero.

To isolate and estimate the uninfluenced behavior we look only at pairs of scores for one person for which the intervening score of their partner was zero (about 15% of the data). That is we are assuming that at such points, $I_{HW} = 0$ and $I_{WH} = 0$, and then Equations 3 and 4 collapse to Equation 2, and we can use least squares on this subset of the data to estimate the two unknown autocorrelation constants for each person. Note that we can now compute the uninfluenced states and inertia of each partner. Now here is an important part of the estimation, the derivation of the influence functions. Once we have estimated the uninfluenced component of the scores we subtract it from the scores at turn t + 1 to find the observed influenced component. We can plot the influenced component of the wife's score against her husband's previous score. This is used to derive one of the influence functions. For each value of the husband's score during the conversation there is likely to be a range of observed values of the influence component due to noise in the data. To convert these into estimates for the influence functions of the model (I_{HW} and I_{WH}), we simply average the observations for each partner score. Both the raw influence data and the averaged influence function are then plotted for each member of each couple.

Steady States and Stability

For each couple, we also plot a phase plane containing the model's null clines. The phase plane refers simply to the plane with the husband's and the wife's scores as coordinates. Hence, a point in this plane is a pair representing the husband's and the wife's scores for a particular interact (a two-turn unit). As time progresses, this point moves, and charts a trajectory in phase space.

We summarize our definitions here. Recall that the *null clines* are the curves in the phase plane for which the derivatives are zero, or the values of the Gottman-Levenson variable stay constant. Once the model has arrived at a stable steady state, it will remain there, hence the term "attractor;" the attractor is similar to a gravitational attractor, it "draws" the values of the model back to it if it is perturbed slightly from the stable steady state. In phase space there are sometimes points called "stable steady states." These are points that the trajectories are drawn toward, and if the system is perturbed away from these states, it will be drawn back. Unstable steady states are the opposite: if perturbed, the system will drift away from these points.

Finding the Null Clines and the Influenced Set Points. It is of considerable importance to find the steady states of the phase plane, the influenced set points, and this is accomplished by finding those points were the null clines intersect. Remember that the null clines are determined by the equations. Finding the null clines is accomplished mathematically by plotting them. Null clines involve searching for steady states in the phase plane; they are theoretical curves where things stay the same over time. A person's null cline is a function of their partner's last score and gives the value of their own score when this is unchanged over one iteration. Mathematically, this is written as:

$$W(t+1) = W(t) = W$$

This last equation says that, for the wife's behavior, things stay the same over time, and that is precisely how we find the shapes of the null clines. Plotting null clines and finding their intersections provides a graphical means of determining steady states. First we begin with simple algebra in which we substitute W for all the wife terms. This process gives us:

$$\begin{aligned} W &= r_1 W + a + I_{HW}(H_t), or \\ W - r_1 W &= I_{HW}(H_t) + a, or \\ (1 - r_1)W &= I_{HW}(H_t) + a, or \\ W &= (I_{HW}(H_t) + a)/(1 - r_1) \end{aligned}$$

That last equation is the wife's null cline. It's the curve where she doesn't change. When we do the same analysis for the husband's null cline, and recall that the steady states are the intersection of the null clines, this then gives the final form of our null clines as:

$$W(H_t) = (I_{HW}(H_t) + a)/1 - r_1)$$
$$H(W_t + 1) = I_{WH}(W_{t+1}) + b/(1 - r_2) \qquad (5)$$

Therefore, we have discovered by simple algebra that our null clines are simply the influence functions, scaled (by 1-r_1, or 1-r_2) and moved (by a or b). In other words, we have shown that the null clines have the same shape as the influence functions, they are moved over (translated) by a constant, and they are scaled by another constant. Null clines play an important role in mathematical analysis since they give a visual indication of the dynamics of the system.

As we noted, the equilibria or steady states are determined by looking for intersections of the null clines, since, by definition, if the system started at this point then it would stay there. Of course, the stability of these steady states to perturbations is yet to be determined. Since we have not specified the functional form of the influence functions, we can only proceed qualitatively.

To derive the influenced steady states of a marital system, the pair of equations (Equation 5) can be solved graphically. The method is similar to solving two simultaneous linear equations ($ax + by = c; dx + ey = f$) by finding intersection points. If these two lines are plotted on the same graph, the point where they intersect gives the solution value (x,y) that satisfies both equations. In our case these functions are not straight lines; they are probably nonlinear (depending on our theory of marital influence). Therefore, if we plot the two curves from Equation 5, their solution will be given by any points where the curves intersect. Now we need to think about what we know about couples' interaction to generate a functional form for the influence functions.

Becoming familiar with influence functions and null clines in the bilinear case. One thing that seems to have emerged from marital research is that, during conflict, negativity has a bigger impact on one's partner's immediately subsequent behavior than positivity. In Figure 2, we depict a graphical summary of this idea and plot both of the bilinear influence functions. Here the husband's influence on his wife is drawn as the dotted line. The solid line is the reverse, the wife's influence function on her husband. This latter function is drawn on the same graph by mentally rotating axes. The positive part of the wife axis, which is vertical, now gets viewed as an abcissa (x-axis), and the first half of the bilinear influence function is then drawn in the positive-wife/positive-husband quadrant. A similar line is drawn in the negative-wife/negative husband quadrant. These influence functions will then be translated and stretched to become the null clines, whose intersection determines the marital system's steady states, given these parameter estimates. Notice that we are plotting two null clines, and doing this graphically is a little tricky. This is illustrated in Figure 3 for the bilinear form of the influence function.

Figure 4 shows how the shape of the null clines at their intersection determines the stability of the steady state. We are plotting two functions: The value of W_t for which $W_{t+1} = W_t$ for any given intervening H_t, and the converse for the husband. Intersection points are, by definition, points for which both the wife's and the husband's score remain constant on consecutive turns of speech. These are the points we call the "influenced steady states." If a couple were to reach one of these states during a conversation, they would theoretically remain there with each partner scoring the same on each of their future turns of speech. If they were perturbed away from one of these steady states, they would be drawn back to it.

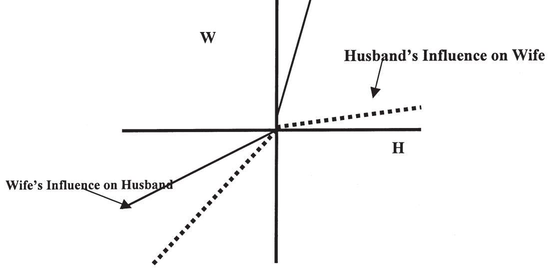

Figure 2. Bilinear form of the influence functions. To draw the wife's influence on the husband (solid line) it is necessary to mentally flip axes, considering the vertical axis as the "wife's abcissa" and the horizontal axis as the "wife's ordinate."

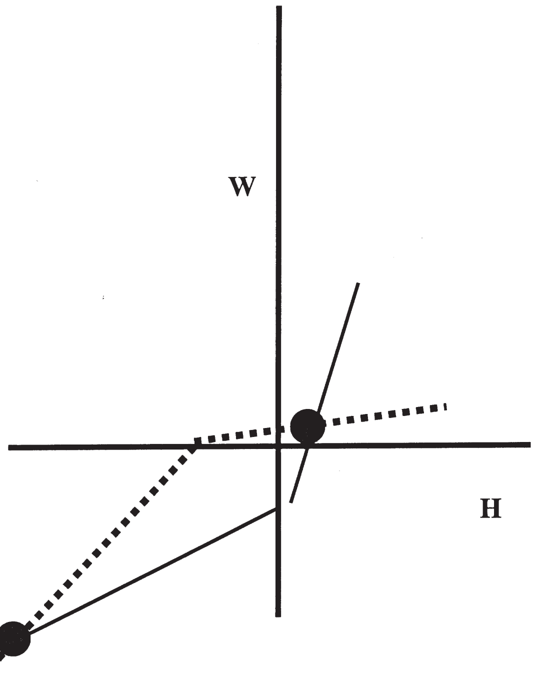

Figure 3. Null clines using the bilinear form of the influence functions. Notice that the null cline for the husband (dotted line) is the husband's influence function on his wife moved to the left (and stretched). The null cline for the wife (solid line) is the wife's influence function on her husband moved down (and stretched). The black dots are the stable steady states.

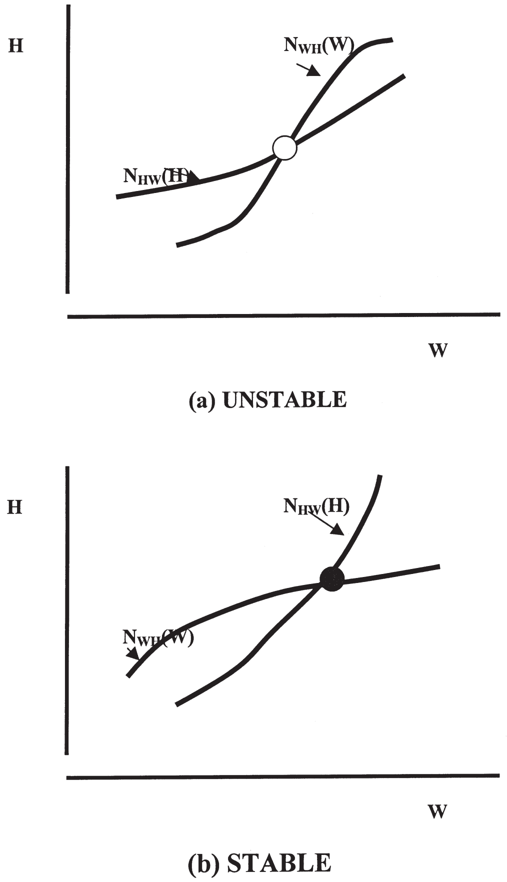

Figure 4. Null clines and the stability of steady states for the marriage model. Steady states correspond to points at which the null clines intersect. The stability of a steady state can be determined graphically (see text for details): when the null clines intersect as is shown in (a), the steady state is unstable; when they intersect as is shown in (b), the steady state is stable.

This perturbation could happen by assuming that there is some random error that also affects people's behavior. In phase space there are paths that each perturbed point will take back to a stable steady state, or away from an unstable steady state. These potential flow lines can be used to map potential trajectories, or solutions to the equations in phase space. Note that theoretically there are two steady states for every couple with the bilinear influence function.

If a couple begins somewhere in state space, all things being equal, it will generally be drawn to the steady state it is closest to. Thus, if a couple begins negatively they are most likely to be drawn to their negative steady state; if a couple begins positively they are most likely to be draw to their positive steady state. However, Murray has shown (Gottman, Murray, Swanson, Tyson, & Swanson, in press) that every attractor in phase space has a strength much like some gravitational fields, which vary with the mass of an object. This strength will determine which attractor is most influential in predicting a couple's conversational trajectory.

Steady States and Trajectories in Phase Space

As we have noted, there are two types of steady states, stable and unstable. Theoretically, if a conversation were continued for a very long time, then pairs of scores would approach a stable steady state and move away from an unstable one. Mathematicians call the set of points that approach a stable steady state (we ignore the possibility of cycles) the "basin of attraction" for that steady state. Theoretically, this very long conversation would be constructed by simply applying Equations 3 and 4 iteratively from some initial pair of scores. The potential existence of multiple stable steady states each with its own basin of attraction has practical implications. The model suggests that the final outcome (positive or negative trend) of a conversation could depend critically on the opening scores of each partner and the strength of each attractor. If the attractors were of equal strength, where one ends up in the phase space is determined by the couple's actual initial conditions. In our experience, we have generally found that the end points can depend critically on starting values.

An observed or a "reconstructed" conversation can be represented in the phase plane as a series of connected points. In addressing the issue of stability of the steady states, we are asking whether the mathematical equations imply that the reconstructed series will approach a given steady state. Analytically, we ask the question of where a couple will move once they are slightly perturbed from their position away from a steady state. The theoretical (stable or unstable) behavior of the model in response to perturbations is only possible once we assume a functional form for the influence functions. For example, for an influence function that has a sigmoidal shape, we can have 1, 3, or 5 steady states, rather than 2.

What does it mean for there to be multiple steady states? It means that these are all possible states for a particular couple. Even if we only observe the couple near one of them in our study, all are possible for this couple, given the equations. Each stable steady state will have a "basin of attraction." This is the set of starting points from which a reconstructed time series will approach the steady state in question. If there is a single steady state, then its basin of attraction is the whole plane; i.e., no matter what the initial scores were, the sequence would approach this one steady state. If, on the other hand, there are two stable steady states (and, necessarily, one unstable one) generally the plane will be divided into two regions (the basins of attraction). If the scores start in the first stable steady state's basin of attraction, then, in time, the sequence of scores will approach that steady state. The same goes for the second steady state and its basin of attraction. The couple begins at the point (W_1, H_1) in phase space, next moves to the point (W_2, H_2), and next moves to the point (W_3,

H_3), and so on, heading for the large black dot that represents the stable steady state intersection of the two null clines.

Notice that this implies that the eventual trend that the conversation follows can be highly dependent on the initial conditions. Thus, high inertia, high influence couples (who are more likely to have multiple steady states) could potentially exhibit a positive conversation on one day and yet not be able to resolve conflict on another. The only difference could be the way the conversation began (their initial scores). The influence functions and uninfluenced parameters would be identical on each day. This discussion makes concrete the general systems theory notion of "first-order" (or more superficial, surface structure) change and "second-order" (or more meaningful, deeper structure) change. In our model, first-order change means that the steady states may change but not the influence functions; second-order change would imply a change in the influence functions as well.

Catastrophes Can Be Represented With This Model

In dynamical theory, "catastrophe" means that a model parameter can change continuously, but, once the parameter crosses a critical threshold, a qualitative change occurs and the same laws no longer govern the system. The classic example is the straw that broke the camel's back. For example, it is possible to change parameters of the marriage model continuously so that the couple loses a positive steady state. This can happen by slowly changing the slopes of the bilinear influence functions, or by slowly changing the parameters a and b. If that happened, suddenly the marital system will have lost its positive influenced steady state. Then no matter where on the basin of attraction the couple started the conversation, they would be inexorably drawn to the negative stable steady state. That would be all that were left to them in phase space! Now, inexplicably all the couple's conflict resolution discussions would degenerate into very aversive and highly negative experiences. This would be a literal catastrophe, and we would predict divorce as inevitable for this couple. This is consistent with Gottman's (1994) report that when that happens, the couple enters a series of cascades that results in increasing flooding, diffuse physiological arousal, arranging their lives in parallel so that they have less interaction, and becoming increasingly lonely and vulnerable to other relationships. To the couple the change is inexplicable. They have weathered many stresses in the past and succeeded in staying together. But now every disagreement heads south toward the negative attractor. That is because there has been a qualitative change: There is no longer a positive attractor.

Here, then, is a model for a very gradual trend during which the couple often thinks that they are simply adapting to increasing stresses in their lives, getting used to seeing less of each other and more fighting, but fully expecting that things will get better eventually. However, they are vulnerable for losing their positive stable influenced steady states, and then the model would predict a real marital catastrophe. The gradual changes would suddenly change the marriage suddenly, and then it would qualitatively become an entirely different relationship.

Although the model predicts this sudden catastrophic change under these conditions, it also predicts what is called "hysterisis" in catastrophe theory, which means that this state of affairs is reversible. Clinical experience suggests that, for longer time spans, this reversibility is probably not the case when the marriage has been neglected for long enough. For example, Buongiorno (1992) reported that the average wait time for couples to obtain professional help for their marriage after they have noticed serious marital problems is about six years. This problem of high delay in seeking help for an ailing marriage is one of the great mysteries in this field of inquiry. It may very well be related to another great mystery, which is the nearly universal "relapse phenomenon" in marital therapy. Recently some of our best scholars (e.g., Jacobson & Addis, 1993) have contended that marital therapy has relapse rates so high (30 to 50 percent within a year after marital therapy ends) that the entire enterprise of marital treatment may be in a state of crisis. Consistent with these conclusions, the recent Consumer Reports study of psychotherapy (Seligman, 1995) reported that marital therapy received the lowest marks from psychotherapy consumers. Marital therapy may be at an impasse because it is not based on a process model derived from prospective longitudinal studies of what real couples do that predicts that their marriages will wind up happy and stable, unhappy and stable, or end in divorce.

After so long a delay before getting help, it makes some sense to propose that a positive hysterisis journey may be less likely than a negative one. Also, some key life transitions may make going back to the more positive way things were less likely. This is particularly true for the transition to parenthood. Half of all the divorces occur in the first seven years of marriage, and a great deal of stress is associated with the transition to parenthood. There are other vulnerable transition points for marriages in the life course. The low point cross-nationally for marital satisfaction is when the first child reaches the age of 14, although this phenomenon is not well understood. Retirement is also such a delicate transition point. If these speculations are true, the model would have to be altered to accommodate these asymmetrical phenomenon.

It does seem likely that there is something like a second law of thermodynamics for marital relationships, that things fall apart unless energy is supplied to keep making the relationship alive and well. At this time in the history of Western civilization, marriages seem more likely to fall apart than to stay together (Martin & Bumpass, 1989). Hence the hysterisis property of the model may turn out to be incorrect. However, our recent research with long-term first marriages (Carstensen, Gottman, & Levenson, 1995) paints a far more optimistic picture, one that suggests that some marriages mellow with age and get better and better.

It should be pointed out that the model is designed to obtain parameters from just a 15-minute interaction, and one useful way of extending the model is to attempt to model two sequential interactions, in which the parameters of the second interaction are affected by the first interaction. What is very interesting about the catastrophic aspects of the model is that it does tend to fit a great deal of our experience, in which we have observed that many marriages suddenly fall apart, often after having successfully endured a period of high stress.

Implications of the model

One of the interesting implications of the mathematical model is that even in the best of marriages, it is possible that there will be both a positive and a negative stable steady state. This means that, depending entirely on starting values, there will be times that the couple will be drawn toward a very negative interaction. This may not happen very often in a satisfying and stable marriage, but the model predicts that it will happen. To some extent, these events are minimized if the strength of the negative steady state (or "attractor") is much smaller than the strength of the positive steady state. This means that the old concept of family homeostasis has to be modified: there are usually at least two homeostatic set points in a family, one more positive than the other. This concept may do a great deal toward ending what Wile (1993) has called the "adversarial" approach of family systems therapy. Here the therapist struggles gallantly against great odds as the family's homeostasis holds on to dysfunctional interaction patterns. However, if there are two homeostatic set points, one more positive and one more negative, then the therapist can align with a family toward making their occupation time greater in the more positive homeostatic set point than in the more negative homeostatic set point.

Another implication of this two homeostatic set point theory is that a negative steady state may have some positive functions in a relationship; the therapist ought not to make war on negative affect, for example. Negativity might be useful in a relationship for a variety of reasons. One is that in any real close relationship, our legacy is the full repertoire of emotions (they are controlled in more formal and more casual relationships). It would not be very intimate if some emotions were expurgated from the full repertoire of emotions that is our legacy as homo sapiens; in fact, a relationship with only positive emotions might actually be a living hell. Second, negative affects may serve the function of culling out behaviors that do not work in the relationship, continually fine tuning the relationship over time so that there is a better and better fit between partners. Third, negativity might serve the function of continually renewing courtship over the course of a long relationship; after the fight there is greater emotional distance, which needs to be healed with a re-courtship. The model has accomplished a great deal just by dismantling the Gottman-Levenson variable into components and parameters. This has created a new theoretical language for describing interaction. Instead of having just a variable that predicts the longitudinal course of marriages, we now can speak theoretically about the mechanism of this prediction. We can expect that compared to happy, stable marriages, what happens in marriages headed for divorce is that:

- there is more emotional inertia;
- even before being influenced, the uninfluenced set point is more negative;
- when interaction begins, the couple influences one another to become even more negative, rather than more positive;
- over time, as these negative interactions continue and become characteristic of the marriage, the couple may catastrophically lose its positive stable steady state.

The model suggests one possible integration of the concepts of affect and power in relationships, which has haunted the field since its inception. The integration is that power or influence is defined as one person's affect having an influence over the other person's immediately following affect. The integration also suggests a greater order of complexity to the concept of power. Who is more powerful in the relationship may be a function of the level of affect, and how positive or negative it is. In one relationship, for example, a wife might be more powerful than her husband only with extreme negative affect, while her husband might be more powerful only with mild positive affect. We may also discover that the very shape of the influence functions are different for couples heading for divorce, compared to happy, stable couples.

Therefore, unlike prior general systems writings, which remained at the level of metaphor, the mathematical model has also given birth to a new theoretical language about the mechanism of change. In the marital research area we did not have such a language before the model was successfully constructed. The model provides the language of set point theory, in which a number of quantities, or parameters, may be regulated and protected by the marital interaction. It also provides a precise mechanism for change. The model itself suggests variables that can be targeted for change using interventions. In short, the model leads somewhere. It helps us raise questions, helps us wonder what the parameters may be related to, and why. It raises questions of etiology. Why might a couple begin an interaction with a negative uninfluenced set point? Why and how would they then influence one another to be even more negative?

Thus, a major contribution of the model is the theoretical language and the mathematical tools it provides. It will give us a way of thinking about marital interaction that we never had before. Previous experience in the biological and physical sciences suggests that any model that accomplishes these things will probably be useful.

The model also gives us insight about its own inadequacy. It is a very grim model. Where a couple begins the interaction in phase space will determine its outcome. Does it make sense to have a model in which there is no possibility of repair? Wouldn't it make more sense to include "switch" terms, which are multiplied by parameters so that when the interaction becomes too negative, the switch is turned on and the interaction gets a jolt of positivity? The parameter could vary for each individual. When we examined our data we found only 4% of the couples who began the interaction negatively but were able to significantly turn the interaction around so it eventually became more positive than negative! Hence, even though major repair is a rare phenomenon, we thought that the model should have such a switch term in it. Couples would then vary in the extent to which they were able to repair the interaction. Perhaps in marital therapy this repair term becomes strengthened. The repair effectiveness parameter could be estimated from small, or "local" turnabouts in the overall direction of the interaction. Also it makes sense to think about the threshold at which repair begins, and whether the repair term is determined by one's own negativity (slope or level) or one's partner's. We have extended the model in this manner. Similarly, once one imagines a term for the down-regulation of negativity, it is possible to have a switch term that down regulates positivity as well as negativity. We have also experimented with this "damping" term in the model. The addition of both a repair and a damping term makes it possible to have many more than two set points for the interaction. Thus, the nature of the nonlinearity determines the complexity of the marital system.

We have also discovered new things with the model, things we never expected to find. In the Cook et al. (1995) article we reported the discovery that different types of marriages have different types of influence functions, and that divorce is only predicted by fundamental mismatches in influence function shape. Cook also showed (Cook et al., 1995) that there is an optimal balance between influence and inertia. Cook's analysis of the stability of the steady states of the marriage model shows that there is a dialectic between the amount of influence each spouse should have on his partner and the level of emotional inertia in each spouse's uninfluenced behavior for the steady state to be stable. Steady states tend to be stable when they have a lower level of influence and a lower level of inertia. Another way of saying this is that if a marriage is going to have high levels of mutual influence, for stability of (say) the positive steady state, there needs to be lowered inertia.

The model continues to develop. For example, it is possible to model not only interactive behavior, but also any time series synchronized to behavior. Hence, we can also model our perceptual (rating dial video recall) data, and the couple's physiology. Indeed, mixed models are possible across measurement domains, and this has led us to propose that in functional marriages there is a "core triad of balance" that is regulated by the couple toward greater positivity in behavior and perception, and greater calm physiologically. In the course of these analyses we discovered that if the wife's behavior drives the husband's physiology, the marriage is likely to end in divorce.

Conclusions

Our work has demonstrated that it is possible to place the study of personal relationships on a solid mathematical footing. We have only begun to explore the utility of this mathematics as we have begun to do proximal change experiments. After pilot testing these proximal change experiments for three years, we have begun a monthly column with the *Reader's Digest* magazine in which once a month a couple comes to our laboratory, we do an assessment (with a pre-intervention conversation that we model) and then perform an intervention, and evaluate the intervention with our mathematical modeling of a post-intervention conversation. At the time of this writing we have successfully worked with 10 couples. These experiments are slowly building a library and a technology of proximal relationship change interventions.

This dream of a mathematics for social relationships is not new. The work we have done is reminiscent of Isaac Asimov's science fiction classic series of books, called the Foundation series. In that series of books, a fictional mathematician named Hari Seldon creates a set of equations for predicting the future of the entire human species, a new branch of study he calls "psychohistory." Considered much harder to accomplish, later in the books another mathematician creates a new set of equations for predicting the future of smaller social units, which he calls "micro-psychohistory." We believe that it is that latter field that we have created. In our view, putting the study of social relationships on a mathematical footing is a major advance in our ability to understand and perhaps to regulate these relationships for the betterment of all mankind.

References

Alexander, J. F. (1973). Defensive and supportive communication in normal and deviant families. *Journal of Marriage and the Family, 35*, 613–617.

Buongiorno, J. (1992). Wait time until professional treatment in marital therapy. Unpublished master's thesis, Catholic University.

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

Carstensen, L. L., Gottman, J. M., & Levenson, R. W. (1995). Emotional behavior in long-term marriage. *Psychology and Aging, 10*(1), 140-149.

Cook, J., Tyson, R., White, J., Rushe, Gottman, J. M., & Murray, J. D. (1995). The mathematics of marital conflict: Qualitative dynamic mathematical modeling of marital interaction. *Journal of Family Pscyhology, 9*, 110–130.

Fincham, F. D., Bradbury, T. N., & Scott, C. K. (1990). Cognition in marriage. In F. D. Fincham & T. N. Bradbury (Eds.), *The psychology of marriage*, (pp. 118–149). New York: Guilford.

Fitzpatrick, M. A. (1988). *Between husbands and wives: Communication in marriage.* Beverly Hills, CA: Sage.

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

Gottman, J. M. (1993). The roles of conflict engagement, escalation, or avoidance in marital interaction: A longitudinal view of five types of couples. *Journal of Consulting and Clinical Psychology, 61*, 6–15.

Gottman, J. M. (1994). *What predicts divorce*? Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Gottman, J. M. (Ed.) (1996). *What predicts divorce: The measures.* New York: Lawrence Erlbaum Associates, Inc.

Gottman, J. M. (1999). *The marriage clinic*. New York: Norton.

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

Gottman, J., Murray, J., Swanson, C., Tyson, R., & Swanson, K. (in press). *The mathematics of marriage: Nonlinear dynamic models.* Cambridge, MA: MIT Press.

Jacobson, N. S., & Addis, M. E. (1993). Research on couple therapy: What do we know? Where are we going? *Journal of Consulting and Clinical Psychology, 61*(1), 85–93.

Maccoby, E. (1990). Gender and relationships: A developmental account. *American Psychologist, 45*(4), 513–520.

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

Margolin, G., & Wampold, B. E. (1981). Sequential analysis of confilct and accord in distressed and nondistressed marital partners. *Journal of Consulting and Clinical Psychology, 49*, 554–567.

Mischel, W. (1968). *Personality and assessment.* New York: Wiley.

Murray, J. D. (1989). *Mathematical biology.* Berlin: Springer-Verlag.

Nye, F. I. (1988). Fifty years of family research, 1937–1987. *Journal of Marriage and the Family, 50,* 305–316.

Patterson, G. R. (1982). *Coercive family process*. Eugene, OR: Castalia.

Revenstorf, D., Vogel, B., Wegener, C., Hahlweg, K., & Schindler, L. (1980). Escalation phenomena in interaction sequences: An empirical comparison of distressed and non-distressed couples. *Behaviour Analysis and Modifications, 2*, 97–ll6.

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

Schaap, C. (1982). *Communication and adjustment in marriage*. The Netherlands: Swets & Zeitlinger B.V.

Seligman, M. E. P. (1995). The effectiveness of psychotherapy: The *Consumer Reports* study. *American Psychologist, 50*, 965–974.

Shannon, C. & Weaver, W. (1949). *The mathematical theory of communication.* Urbana: University of Illinois Press.

Terman, L. M., Buttenweiser, P., Ferguson, L. W., Johnson, W. B., & Wilson, D. P. (1938). *Psychological factors in marital happiness*. New York: McGraw Hill.

Ting-Toomey, S. (1982). An analysis of verbal communication patterns in high and low marital adjustment groups. *Human Communication Research, 9*, 306–3l9.

von Bertalanffy, L. (1968). *General system theory.* New York: Braziller.

VonNeumann, J., & Morgenstern, O. (1947). *Theory of Games and Economic Behavior.* Princeton, NJ: University of Princeton Press.

Wiener, N. (1948). *Cybernetics*. New York: Wiley.

Wiggins, J. S. (1973). *Personality and prediction.* Reading, MA: Addison-Wesley.

Wile, D. B. (1993). *After the fight: A night in the life of a couple*. New York: Guilford Press.