Journal of Family Psychology
1995, Vol. 9, No. 2, 110–130

0893-3200/95/$3.00

Mathematics of Marital Conflict: Qualitative Dynamic Mathematical Modeling of Marital Interaction

Julian Cook, Rebecca Tyson, Jane White, Regina Rushe, John Gottman, and James Murray
University of Washington

This article presents a new nonstatistical mathematical approach to modeling marital interaction by means of nonlinear difference equations. The application made of this modeling process in this article is to generate theory in order to understand J. M. Gottman and R. W. Levenson's (1992) report of the ability of one variable derived from a balance view of marital interaction to predict marital dissolution in a prospective study. Parameters are introduced that reflect uninfluenced husband and wife set points, emotional inertia, and influence function. These parameters are subjected to various examinations of validity with other data that reflect patterns of marital interaction and the cascade toward marital dissolution. Results suggest the hypothesis that unstable marriages are characterized at Time 1 by a mismatch in husband and wife influence functions. The modeling method makes it possible to fit a couple's equations under one set of conditions, then to simulate the couple's interaction under different conditions, and then to conduct experiments to test the validity of these simulations. The method provides a new approach toward the building of theory in family psychology.

This article introduces a new approach to the modeling of 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; rather they are designed to suggest a precise mechanism of change. This method has been used with great success in the biological sciences (Murray, 1989). The method is usually a quantitative approach that requires the modeler to be able to write down in mathematical form, on the basis of some theory, the causes of change in the dependent variables. For example, in the classic predator–prey problem, a modeler writes down that the rate of change in the population densities is some function of the current densities (e.g., Murray, 1989). Although this example is a simple representation of the predator–prey phenomenon, it has served well as an initial exploratory model. The equations are designed to enable modelers 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 use linear terms or linear approximations of nonlinear terms, and they often give very good results. In fact, most of the statistics used in the field of family psychology are based on 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]

However, in recent years it has become clear that most systems are complex and must be described by nonlinear terms. It is interesting that by the use of nonlinear terms in equations of change, some very complex processes can be

Julian Cook, Rebecca Tyson, Jane White, and James Murray, Department of Applied Mathematics, University of Washington; Regina Rushe and John Gottman, Department of Psychology, University of Washington.

Correspondence concerning this article should be addressed to John Gottman, Department of Psychology, NI-25, University of Washington, Seattle, Washington 98195.

[1] For example, the equation $y' = 2y$ is linear, whereas the equation $y' = 2\sin(y)$ is nonlinear. For regions of y close to zero, however, $\sin(y) = y$ is a good approximation, and so the linear equation is a good approximation of the nonlinear when y is close to zero.

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 are called *qualitative,* and visual methods must be relied on. 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 in engaging the intuition of a scientist working in a field that has no mathematically stated theory. If a 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.

In modeling marital interaction, we confronted an interesting dilemma. 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. Therefore, we developed instead 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 our 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.

It is our belief that the "test" of these qualitative forms of change should not be an automatic process, as with a *t* test. Instead, we suggest that the data be used to guide scientific intuition so that equations of change are theoretically meaningful. It is this use of mathematical modeling, which generated a theory of change in marriages, that we explore in this article. In an area where it is difficult to use a priori quantitative mathematical theory for describing the processes of interaction, we have found it useful to use a qualitative mathematical modeling approach, the purpose of which is the generation of mathematical theory. We believe that this approach is valuable and quite general. Why would one wish to pursue this kind of mathematical modeling at all? The answer is that it can provide two new things: First, the modeling provides a new language for thinking about marital interaction and change over time, and second, once equations are compiled for a couple, their behavior can be simulated in circumstances other than those that generated the data. Precise experiments can then be done to test whether these simulations are valid. In this manner, theory is built and tested through the modeling.

In our case, we began with a phenomenon, recently reported by Gottman and Levenson (1992), that one variable descriptive of specific interaction patterns of the balance between negativity and positivity was predictive of marital dissolution. We set out to try to generate theory that might explain this phenomenon.

Method

Review of Gottman and Levenson

Gottman and Levenson (1992) used a methodology for obtaining synchronized physiological, behavioral, and self-report data in a sample of 73 couples who were followed longitudinally between 1983 and 1987. By means of observational coding of interactive behavior with the Rapid Couples Interaction Scoring System (RCISS; Krokoff, Gottman, & Hass, 1989), couples were divided into two groups, called *regulated* and *nonregulated.* 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.[2] 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. Then, the cumulative total of these points was plotted for each spouse (see Figure 1). 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 through linear regression analysis. If both husband and wife graphs had a positive slope, they were called *regulated;* if not, they were called *nonregulated.* This classification is referred to as the *Gottman–Levenson variable.* Using Cohen's kappa, reliability for all RCISS subcodes taken together was .72. All couples, even happily married ones, had some amount of

[2] These codes were combined into the following four subscales for each spouse: (a) Complaint/Criticism, (b) Defensiveness, (c) Contempt, and (d) Listener Withdrawal From Interaction. For the individual speaker codes, kappas ranged from .70 to .81.

negative interaction; similarly, all couples, even unhappily married ones, had some degree of positive interaction. 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. Thus, the terms *regulated* and *nonregulated* have a very precise meaning here. Regulated couples were defined as those for whom both husband and wife speaker slopes were significantly positive; nonregulated couples had at least one of the speaker slopes that was not significantly positive. By definition, regulated 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 regulated couples and 31 nonregulated couples.[3]

1987 Follow-Up

In 1987, 4 years after the initial assessment, the original participants were recontacted and at least one spouse (70 husbands and 72 wives) from 73 of the original 79 couples (92.4%) agreed to participate in the follow-up. Information on marital status was obtained.

Observational Coding: Validity Measures

The videotapes of the problem area interaction were coded with the following two observational coding systems: The RCISS provided the means for classifying couples into the regulated and nonregulated marital types, as well as for providing base rates of specific positive and negative speaker and listener codes. The Marital Interaction Coding System (MICS; Weiss & Summers, 1983) was used as measures of convergent validity. MICS codes were collapsed into three negative summary speaker codes: (a) defensiveness, which was the sum of excuse, deny responsibility, negative solution, and negative mind reading by the partner, (b) criticism, which was the sum of disagreement and criticism, and (c) contempt, which was the sum of noncompliance, verbal contempt, command, and complaint. Despite the fact that we were to model the speaker's behavior, we also included a fourth summary MICS code that describes the listener's disengagement from the interaction: withdrawal from interaction, which is the sum of negative listener behaviors, no response, not tracking, turn off, and incoherent talk. Codes were assigned continuously by coders for 30-s blocks.[4] Means reported for the MICS are based on the total number of codes in 15-min periods. A sample of every videotape was independently coded by another observer and a confusion matrix (i.e., a matrix of counts of agreements and disagreements for two observers) for each code category was computed. The average weighted Cohen's kappa for this coding (all individual subcodes, summed over all couples) was .60. For the four negative summary codes, the overall kappas were higher, ranging between .65 and .75.

Positive Affect

One of the first things to disappear when a marriage is ailing is positive affect, particularly humor and smiling. In this study, the parameters of our equations were also correlated with the amount of laughter (assessed with the RCISS) and the amount of smiling (measured by coding facial expressions with Ekman and Friesen's, 1978, Facial Action Coding System). Only Duchenne smiles (which include both zygomatic and orbicularis oculi contraction) were measured, since these have been found to be related to genuine felt positive affect.

Results From Gottman and Levenson

Figure 2 summarizes the Gottman and Levenson (1992) 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.

Marital Typology

Gottman (1994) proposed and validated a typology of three types of longitudinally stable marriages with distinct Time 1 marital interaction patterns; these interaction patterns differed from the Time 1 interaction patterns of couples heading for dissolution. There were three groups of stable couples: validators, volatiles, and avoiders, who could be distinguished on problem-solving behavior and specific affects and, through the use of log-linear analysis, on one variable designed to provide an index of the amount and timing of persuasion attempts. There were two groups of unstable couples: hostile and hostile–detached, who could be distinguished from one another on problem-solving behavior and on specific negative and positive affects. The hostile–detached group was significantly more negative (more defensive and contemptuous) than the hostile group. Gottman (1993) reported that there was a rough constant that was invariant across each of the three types of stable

[3] We model the unaccumulated data later in this article.

[4] The MICS currently employs *double codes,* which refer to coding a behavior as simultaneously both one code and another code. Double codes were treated as additional single codes for this research.

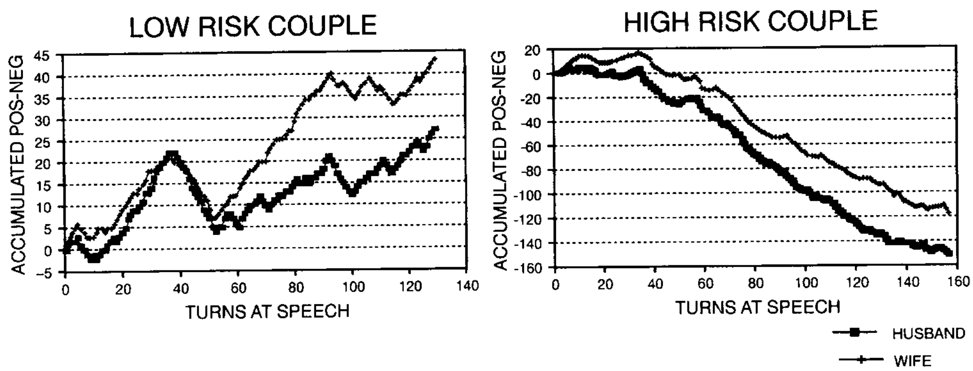

Figure 1. Cumulative Rapid Couples Interaction Scoring System (RCISS) speaker point graphs for a regulated (low risk) and a nonregulated (high risk) couple. Pos-Neg = Positive–Negative. Revised art from "Marital Processes Predictive of Later Dissolution: Behavior, Physiology, and Health," by J. M. Gottman and R. W. Levenson, 1992, *Journal of Personality and Social Psychology, 63,* p. 225. Copyright 1992 by the American Psychological Association.

couples. This constant, the ratio of positive to negative RCISS speaker codes during conflict resolution, was about 5, and it was not significantly different across the three types of stable marriages. Perhaps each adaptation to achieve a stable marriages, or each stable couple type represents a similar kind of adaptation, although the marriages were quite different. The volatile couples reached the ratio of 5 by mixing a lot of positive affect with a lot of negative affect. The validators mixed a moderate amount of positive affect with a moderate amount of negative affect. The avoiders mixed a small amount of positive affect with a small amount of negative affect. Each does so in a way that achieves roughly the same balance between positive and negative. We can speculate that each type of marriage has its risks, benefits, and costs. It is possible to speculate about these risks, costs, and benefits based on what we know about each type of

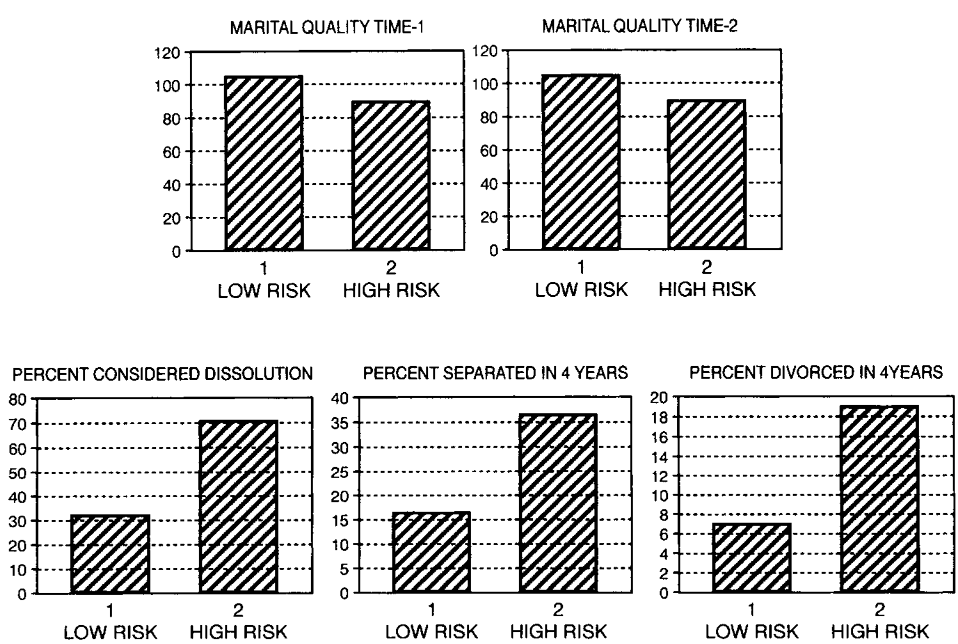

Figure 2. Outcome data from the marital dissolution cascade for high- and low-risk marriages.

marriage. The volatile marriage tends to be quite romantic and passionate, but has the risk of dissolving to endless bickering. The validating marriage, which is the current model used in marital therapy, is calmer and intimate; these couples appear to place a high degree of value on companionate marriage and shared experiences, not on individuality. The risk may be that romance will disappear over time and that the couple will become merely close friends. Couples in the avoiding marriage avoid the pain of confrontation and conflict, but they risk emotional distance and loneliness. Gottman (1994) also found that the three types of stable marriages differed in the amount and timing of persuasion attempts. Volatile couples engaged in high levels of persuasion and did so at the very outset of the discussion. Validators engaged in less persuasion than did volatile couples and waited to begin their persuasion attempts until after the first third of the interaction. Conflict-avoiding couples hardly ever attempted to persuade one another. We wondered whether these five types of marriage could be discriminated using the parameters and functions derived from the mathematical modeling.

Mathematical Modeling

The goal of the mathematical modeling was to dismantle the RCISS point graphs of (unaccumulated) positive minus negative behaviors at each turn into components that had theoretical meaning. This process was an attempt at understanding the ability of these data to predict marital dissolution through the interactional dynamics. We began with the Gottman–Levenson dependent variable and dismantled it into components that represented (a) a function of interpersonal influence from spouse to spouse and (b) the terms containing parameters related to an individual's own dynamics. This dismantling of RCISS scores into influenced and uninfluenced behavior represents 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 portion of 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. To illustrate the selection of an analytical form for the influence function, we can begin with the simple assumption that there is a threshold before a positive value has an effect in a positive direction and another threshold before a negative value has an effect in a negative direction. A more reactive spouse has a lower threshold of response. The parameters of these influence functions (e.g., the point at which the spouse's negativity starts having an effect) might vary as a function of culture, marital satisfaction, level of stress the spouses were under at the time, their individual temperaments, and so forth. These latter ideas can be used at a later time to improve the model's generality and predictive ability. We then assume that the amount of influence will remain constant across the remainder of the ranges of the variable. This is, of course, only one kind of influence function that we could have proposed. For example, we could have proposed that the more negative the dependent variable, the more negative the influence, and the more positive the dependent variable, the more positive the influence (reminiscent of Alexander's defensive–supportive cycle; e.g., Alexander, 1973). The two options are depicted in Figure 3. The top half of the figure shows an influence function that remains constant once there is an effect (either positive or negative), and the bottom half shows an influence function in which the more positive the previous behavior, the more positive the effect on the spouse, and the more negative the behavior, the more negative the effect on the spouse.

We began with a sequence of RCISS scores: W_t, H_t, W_{t+1}, H_{t+1}, . . . etc. In the process of modeling, two parameters are obtained for each spouse. One parameter is their emotional inertia (positive or negative), which is their tendency of remaining in the same state for a period of time, and the other is their natural uninfluenced set point, which is their average level of positive minus negative scores when their spouse's score was zero, that is, equally positive and negative.[5] For purposes of estimation, we assumed that zero scores had no influence on the partner's subsequent score. Having estimated these parameters from a subset of the data, we then subtracted the uninfluenced effects from the entire time series to reveal the influence function, which summarizes the partner's influence. An additional parameter that emerged from our modeling was the influenced set point of the interaction, which is a *steady state,* or a sequence of two scores (one for each partner) that

[5] This uninfluenced set point need not be viewed as an individual variable, such as the person's mood or temperament. It could be thought of as the cumulated effect of both the marriage up to the time of observation as well as any propensities this individual has to act positively or negatively at this time. Thus, if a second interaction is observed (particularly following an intervention), it might be of some interest to attempt to predict changes in this parameter over time. It might be also of interest to determine the stability of a person's uninfluenced set point across other relationships, for example comparing marital, parent–child, or friendship interactions.

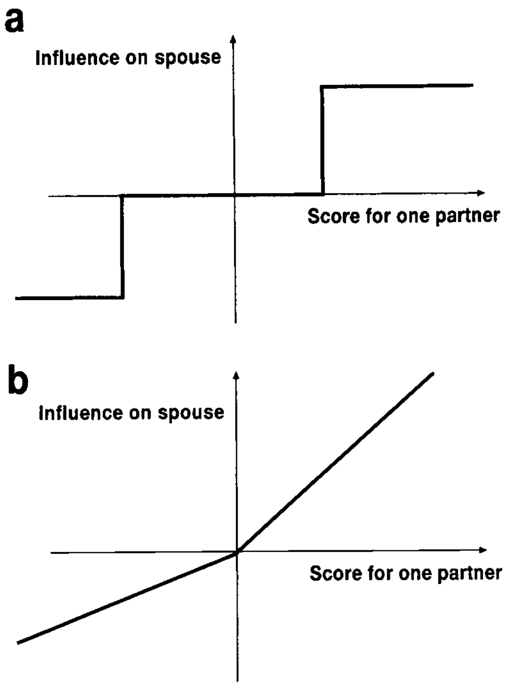

Figure 3. Two possible functional forms for the influence functions. For the influence function of the husband on the wife, the *x*-axis is the husband's previous score, H_t, and the *y*-axis is the influenced component, $I_{HW}(H_{t+1})$, of the wife's following score, W_{t+1}. The wife's influence on the husband, $I_{WH}(W_{t+1})$, could be graphed in a similar way. In Part a, there is no influence unless the partner's previous score lies outside some range. Outside that range the influence takes either a fixed positive value or a fixed negative value. In Part b, influence increases linearly with the value of the previous score, but negative scores can have either a stronger or less strong influence than positive scores. In both graphs, a score of zero has zero influence on the partner's next score (one of the assumptions of the model).

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, the steady state represents the average score the theoretical model would predict for each partner. We thought it might be interesting to examine whether the influenced set point was more positive than the uninfluenced set point—that is, did the marital interaction pull the individual in a more positive or a more negative direction?

We tested the validity of the two parameters estimated for each spouse in each couple by correlating these parameters with the dissolution cascade variables, the behavioral observation scales of the MICS and the RCISS, and our positive affect indices (laughter and Duchenne smiles).

Model-Building Strategy

Our model-building strategy follows the philosophy of Murray (1989). In the book *Mathematical Biology,* Murray constructed fairly simple nonlinear models for complex biological problems. The strategy of model construction is first to propose equations that are good, but simple, representations of the underlying biology. Subsequently, the models and their qualitative solutions are extended by other factors. Hence, we too began simply by modeling marital interaction. We expect to extend out equations by suggesting later that some of our parameters may not actually be fixed constants, but may vary with other variables in the experiment.

The Model

The model presented in this article attempts to reproduce the sequence of RCISS speaker scores. For the present, we confined ourselves to a deterministic model, regarding any score as being determined only by the most recent two scores. In this way, we used a discrete model to describe the individual's level of positivity in each turn at speech. That is, we sought to understand interactions as if individual behavior were based purely on predefined reactions to (and interpretations of) recent actions (one's own and one's partner's). This scenario may not be true in the main, but it may be true enough that the results of the model would then suggest underlying patterns that affect the way any particular couple interacts when trying to resolve conflict.[6] In the next section, the details of the model are described, and in the following section, the methods for estimation of model parameters are described. In the subsequent section, the mathematical and the experimental results are presented.

Description of the model. The assumption that each person's score is determined solely by each person's own and the partner's previous score restricted us to a particular class of mathematical models. If we denote W_t and H_t as the husband's and wife's scores, respectively, at turn *t*, then the

[6] The form of the model is in marked contrast to game theory models, in which there is a presumed matrix of rewards and costs and a goal of optimizing some value. We posit no explicit optimization or individual goal. Each individual simply has a natural state of positivity or negativity and an inertia (related to how quickly displacements from the natural state are damped out), on top of which the partner's influences and random factors act. We do not introduce the concept of a "strategy."

sequence of scores is given by an alternating pair of coupled difference equations:

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

The functions f and g remain to be determined. The asymmetry in the indices is due to the fact that we assumed, without loss of generality, that the wife speaks first. We therefore labeled the turns of speech $W_1, H_1, W_2, H_2, \ldots$. To select a reasonable f and g, we made some simplifying assumptions. First, we assumed 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 each person's previous score only and the other of which depends on the score for the partner's last turn of speech. We termed 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. 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 tend to be more negative when left to themselves while others would naturally be more positive in the same situation. We termed this baseline temperament as the individual's *uninfluenced set point.* We supposed that individuals 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 led us to the linear relationship

$$P_{t+1} = r_i P_t + a_i, \tag{2}$$

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 henceforth is referred to as *inertia.* The uninfluenced set point is the steady state of this equation and is found by solving $P_{t+1} = P_t = P = a_i/(1 - r_i)$. The behavior of this difference equation is governed by the value of r_i. If the absolute value of r_i is less than one, then the system will tend toward the steady state regardless of the initial conditions, whereas if the absolute value of r_i is greater than one, the system will always evolve away from a steady state.

Clearly, the natural state needs to be stable, so we were only interested in the case in which the absolute value of r_i was less than one. The magnitude of r_i determines how quickly the uninfluenced state is reached from some other state or how easily people change their frame of mind, hence the use of the word *inertia.* For selecting the form of the influenced component of behavior, various approaches can be taken. 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. The first approach is to write down a theoretical form for these influence functions (recall Figure 3). For example, we can posit a two-slope function: Two straight lines go through the origin with two different slopes, one for the positive range and one for the negative range. Another possible function that made sense to us was sigmoidal, or S shaped. In this function, again around zero on the x-axis, there is no influence; there is an influence only after some threshold in positivity is passed, and then the influence is positive and constant throughout the positive ranges. Similarly, as a threshold in negativity is passed, the influence is negative and then constant throughout the negative range. Note that other forms of the influence function are also reasonable. For example, an investigator could combine the two functions and have a threshold and two slopes: We simply assumed that there were slopes for negative and positive influences only after the thresholds are exceeded. How are researchers to be guided in the choice of a theoretical influence function? We suggest beginning very simply, with a function that requires very few parameters. The model can be made more complex later, once this complexity is shown to be necessary. In this article, we discuss both the two-slope and the sigmoidal functions.

An alternative approach to the selection of influence functions is to make no attempt to predetermine the form of the function. We did not follow this approach. Instead, we expected the influence functions to vary from person to person, and we decided that one of the aims of our model building at this stage in the research was to uncover the shape of the influence function from the data. Hence, we decided on this approach. In the first study in which we are building the mathematical model, we proceeded entirely empirically. However, we use the data to reveal the influence functions. We will summarize these results using the two-slope form of the influence function. This means that the goal of our mathematical modeling at this point is to generate theory. We denote the influence functions by $I_{AB}(A_t)$, or the influence of person A's state at turn t on person B's state. With these assumptions the complete model is

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

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

Again, the asymmetry in the indices 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.

Estimation of parameters and the unknown influence functions. To isolate and estimate the uninfluenced behavior, we looked only at pairs of scores for one person for which the intervening score of the partner was zero (about 15% of the data). Since, by assumption, $I_{HW} = 0$ and $I_{WH} = 0$, 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 constants for each person. Note that we can now compute the uninfluenced states and inertia of each partner.[7]

Once we have estimated the uninfluenced component of the scores, we can subtract it from the scores at turn $t + 1$ to find the observed influenced component. We can, for example, plot the influenced component of the wife's score against her husband's previous score. 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 that is 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 averaged the observations for each partner score. Both the raw influence data and the averaged influence function can be plotted for each member of each couple.

To validate the estimation process, we then formed a reconstructed conversation from the model equations. We started by simply taking both people to be at their uninfluenced state (noninteger values are allowed in this reconstruction) and then iterated forward for the approximately 80 turns of speech each we tended to observe in 15 min. This was done by computing the components separately and then summing to generate the next score. The uninfluenced component was derived from the use of Equation 2. The influenced behavior was computed by simply rounding the partner's last score to the nearest integer and reading off the influence from that person's average influence function, referred to above. The reconstructed conversation, therefore, lacks any randomness. We do not pretend that this "expected" conversation would ever be observed in practice. Rather, it represents an underlying trend.

Steady states and stability. For each couple, we plotted a phase plane containing the model's null clines. The *phase plane* refers 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. In phase space there are sometimes points called *stable steady states,* which are points that the trajectories are drawn toward. 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. Hence, it is of considerable importance to find the steady states of the phase plane. This procedure is accomplished mathematically by plotting the null clines. Null clines involve a search 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 the partner's last score and gives the value of that person's own score when this is unchanged over one iteration, or $W(t + 1) = W(t)$. This last equation says that things stay the same over time, and that is precisely how we find the shapes of the null clines. Plotting null clines provides a graphical means of determining steady states. Simple algebra gives the form of these null clines as

$$\begin{aligned} W(H_t) &= [I_{HW}(H_t) + a]/(1 - \mathrm{r}_1) \\ H(W_{t+1}) &= [I_{WH}(W_{t+1}) + \mathrm{b}]/(1 - \mathrm{r}_2). \end{aligned} \tag{5}$$

Notice that these equations are simply the influence functions, scaled (by $1 - \mathrm{r}_1$ or $1 - \mathrm{r}_2$) and translated (by a or b). In other words, 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 often play an important role in mathematical analysis since they give a visual indication of the dynamics of the system. 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. Because we have not specified the functional form of the influence functions, we can only proceed qualitatively. However, it is instructive to discuss what would happen if we had settled on a functional form for the influence functions (which we will do in subsequent studies as we develop the model). For example, suppose we assume the sigmoidal S-shaped form. This assumption is reasonable because it presumes two thresholds of influence, and that the influence is bounded in both negative and positive ranges. The pair of equations (Equation 5) can be solved graphically. The method is identical to solving two simultaneous linear equations ($\mathrm{a}x + \mathrm{b}y = \mathrm{c}$; $\mathrm{d}x + \mathrm{e}y = \mathrm{f}$). If these two lines are plotted on the same graph, the point at which they intersect gives the solution value (x,y) that satisfies both equations.

[7] Note that if these zero points were rare, it would be hard to obtain accurate estimates for the model parameters as the confidence intervals around these parameters would be large. Although it seems like a strong assumption, the assumption that zero scores have zero influence is arbitrary. We could have assumed nonzero influences, made these additional parameters, and estimated these parameters as well. In fact, an asymmetry in these parameters would be theoretically interesting in characterizing a couple's interaction. In the interest of parsimony, we picked zero for these parameters.

Therefore, if we plot the two curves from Equation 5, their solution is given by any points where the curves intersect. Call one of the partners A. Under what conditions will two consecutive scores for A be identical? For any particular score for A, there is only one intervening score for A's partner that allows this. Thus, the map from A's score to the partner's score that leaves it unaltered defines a function. This is what we have called A's *null cline.*

A's partner has a null cline that can be computed in a similar way. We plotted the wife's and husband's null clines against their corresponding axes in Figure 4. This plot represents 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 scores remain constant on consecutive turns of speech. We call these points 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 stable steady states, they would be drawn back to it. These potential flow lines can be used to map potential trajectories, or solutions to the equations in phase space. Although there may be many influenced steady states (depending on the influence functions and the uninfluenced parameters), in practice we have usually found only one.

There are actually two types of steady states, stable and unstable. If a theoretical conversation were continued for some time, then pairs of scores would approach a stable steady state and move away from an unstable one. We 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. An example of a sequence of scores is shown in Figure 5 approaching the more positive steady state. This *theoretical 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.

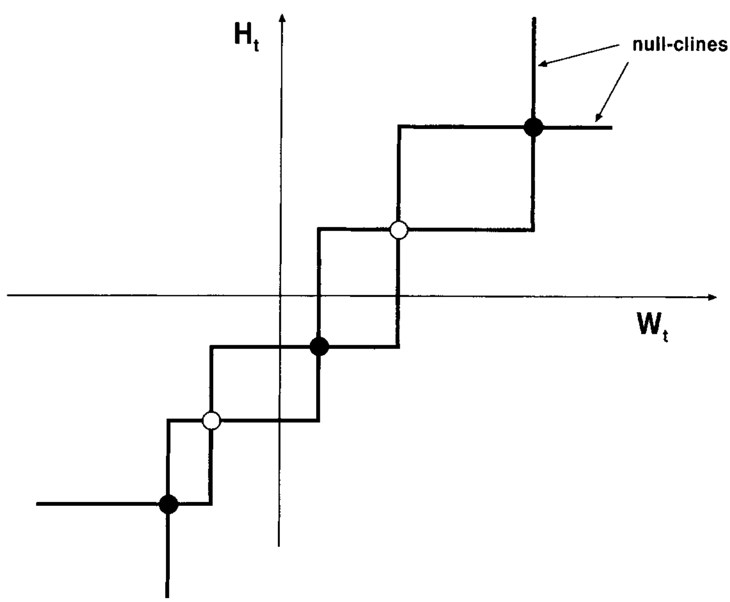

Figure 4. The use of null clines to graphically determine the steady states. The husband's null cline is a function of the wife's previous score. For a particular value of the wife's (W) score, W_{t+1}, the value of the husband's null cline is the value of the husband's (H) score that would remain unchanged over one iteration ($H_t = H_{t+1}$) when the wife's intervening score was W_{t+1}. The wife's null cline is defined in a similar way. The null clines have the same general shape as the influence function (see Equation 5) but are stretched and translated. When the null clines are plotted against their respective axes, the steady states of the system are the points of intersection. The intersection of the two null clines shows that five stability points (one, three, or five stability points) are possible *with this sigmoid form of the influence function;* note that the influence functions are plotted in a mirror-image relationship. The example shown has five stability points. Notice that the stable steady states (filled circles) alternate with unstable steady states.

The model suggests that the final outcome (positive or negative trend) of a conversation could depend critically on the opening scores of each partner. Where one begins in the phase space is determined by the couple's actual initial conditions. We have generally found that the end points can depend critically on starting values.[8] 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 steady state will move once it is slightly perturbed from its position. The theoretical (stable or unstable) behavior of the model in response to perturbations of the steady states is only possible once we assume a functional form for the influence functions. For example, as we have noted, for the sigmoidal influence function, we can have one, three, or five steady states (see Figure 4). From the null cline plot (see Figure 4), we can see that there are three stable and two unstable states.

What does it mean for there to be multiple steady states? These states are all possible for a particular couple. Even if we only observe the couple near one of the steady states in our study, all are possible for this couple, given the equations. Each stable steady state will have a *basin of attraction,* which 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—that is, no matter what the initial scores were, the sequence would approach this one steady state. We have found this tendency toward a single steady state to be the usual situation in our data. 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; see Figure 5). 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. This situation is depicted in Figure 5. 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 forth heading for the large black dot that represents the stable steady state intersection of the two null clines. Notice that this movement implies that the eventual trend which 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 RCISS 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 implies a change in the influence functions as well.

Results

Influence Functions

Notice that influence functions are arbitrarily attributed to the influencer, although we recognize that the influenced spouse also plays a part in determining the influence. As a rough approximation to the shape of the influence functions, obtained from the data by least squares, we used the two-slope function and computed the slope of the influence function separately for negative and positive value of the partner's behavior. The x-axis represented the range of positivity or negativity in each group. Only data close to the natural set point for each group could be trusted to avoid infrequent numbers of instances of RCISS values within a group. Note that this means that we get more reliable information for regulated couples in the positive ranges and for nonregulated couples in the negative ranges of the x-axis. Figure 6 is a summary of the empirically obtained functions for five groups of couples, the three stable marriages (volatile, validating, and avoiding) and the two unstable marriages (hostile and hostile–detached). For heuristic purposes, we used the two-slope model of the influence function. The top three rows of graphs represent the influence

[8] Notwithstanding what has been termed the *punctuation fallacy,* in which where one starts in an interaction is quite arbitrary, we have found that the couple's starting values of the interaction appear to be very important in determining the couple's eventual trajectory. We have considered modifying the influence functions to include a *repair component,* whose existence would be capable of moving a couple from a negative to a positive steady state. If there were a repair component operating, the cumulative graph could look like a check mark, starting downward and then changing direction. Unfortunately, this occurred in our data for only 4% of the cases. Perhaps effective marital therapy might add such a repair component to the influence functions.

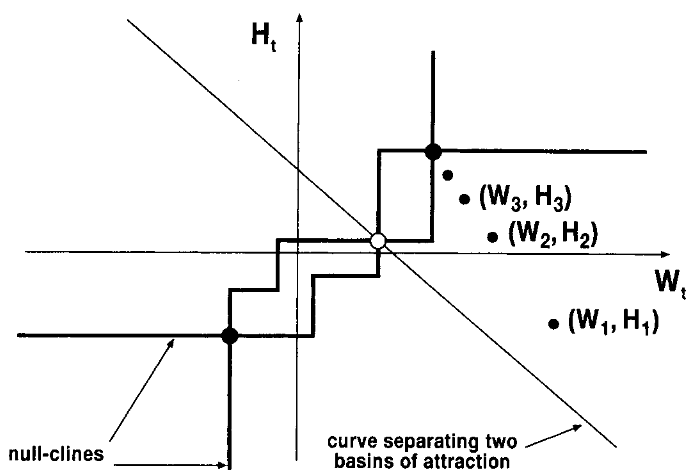

Figure 5. Null clines and the sequence of theoretical Rapid Couples Interaction Scoring System (RCISS) scores in the case of three steady states. Two stable steady states (large filled circles) are separated by an unstable steady state (large open circle). A possible sequence of scores is shown approaching the positive steady state (small filled circles). Each point corresponds to a consecutive pair of scores (W_t, H_t). Both stable steady states have a basin of attraction consisting of points starting from which a sequence of theoretical RCISS scores will approach the steady state in question. The curve separating the basins of attraction is shown as a straight line. Pairs of scores gradually approach one of the stable states: The long-term behavior of the sequence is therefore dependent on the basin of attraction in which the initial pair of scores lies. H = Husband; W = Wife; subscript t = time.

functions for the three regulated marriages. The validators have an influence function that creates an influence toward negativity in a spouse if the partner's behavior is negative, and an influence toward positivity if the partner's behavior is positive. Volatile and conflict-avoider influence functions appear to be, respectively, one half of the validators', with volatiles having the right half of the curve with a slope close to zero and the conflict avoiders having the left half with a slope near zero. This observation of matching functions is summarized in the third column, labeled *theoretical influence function.* Now examine the influence functions for the hostile and the hostile–detached couples. It looks as if these data would support a mismatch hypothesis. Hostile couples appear to have mixed a validator husband influence function with a conflict-avoider wife influence function, and hostile–detached couples appear to have mixed a husband validator influence function with a volatile wife influence function.

From examining the data, we propose that validating couples were able to influence their spouses with either positive or negative behavior; positive behavior had a positive sloping influence, while negative behavior also had a positive sloping influence. This result means that the negative *x*-axis values had a negative influence, whereas the positive *x*-axis values had a positive influence. For validators, across the whole range of RCISS point values, the slope of the influence function was a constant, upwardly sloping straight line. The data might have been generated by the process that in validating regulated marriages there is a uniform slope of the influence function across both positive and negative values: Overall negative behavior has a negative influence, whereas positive behavior has a positive influence in regulated marriages. Here we see that a full range of emotional balance is possible in the interaction. However, conflict-avoiding and volatile couples were nearly opposite in the shape of their influence functions. Avoiders influenced one another only with positivity (the slope was flat in the negative RCISS point ranges), whereas volatile couples influenced one another primarily with negativity (the slope was flat in the positive RCISS point ranges). The

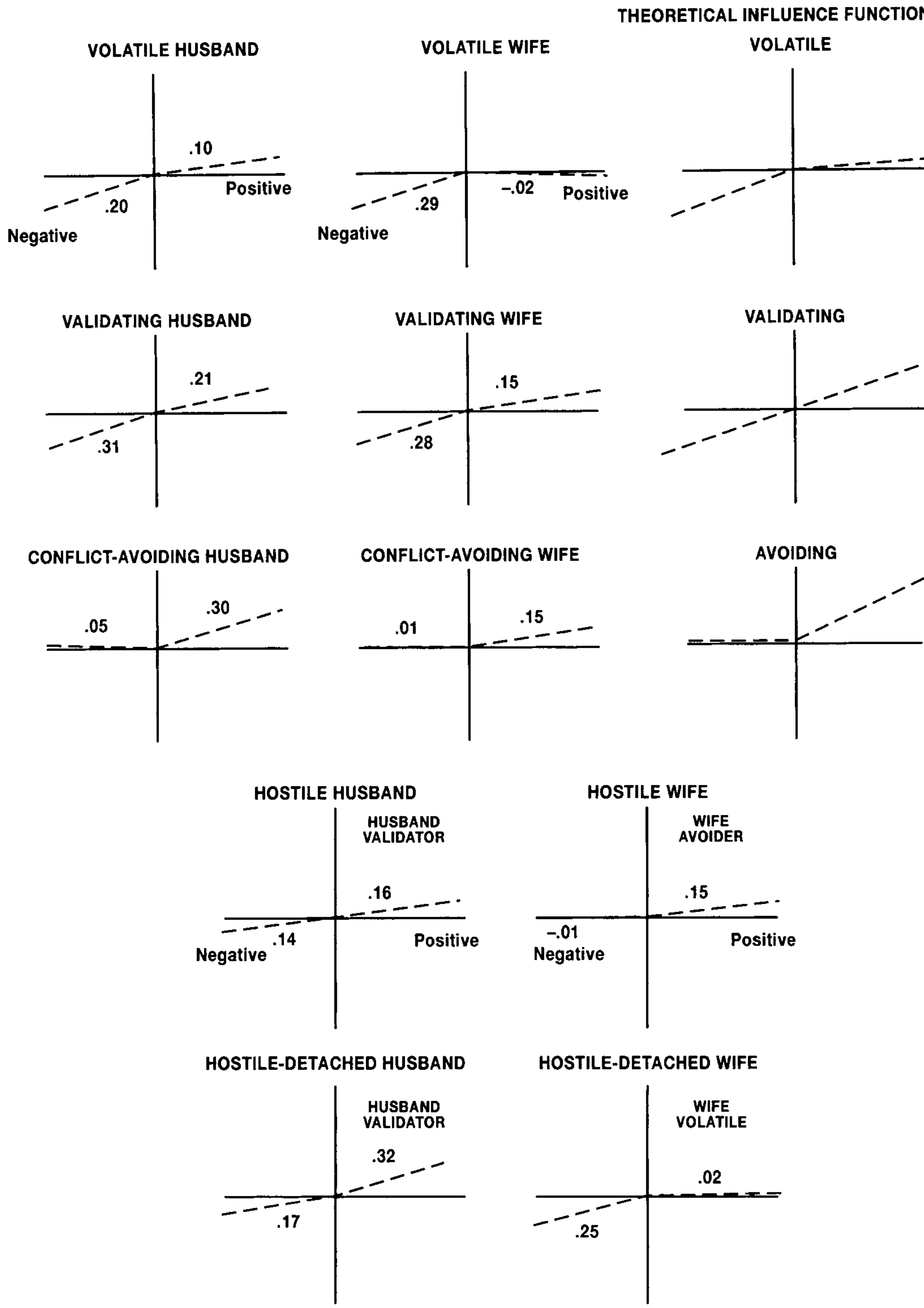

Figure 6. Empirically obtained influence functions using the two-slope functional form.

influence function of the avoiding couple is nearly the reverse of that of the volatile couple.

Mismatch Theory: The Possibility That Unstable Marriages Are the Results of Failed Attempts at Creating a Pure Type

The shape of the influence curves leads us to propose that the data on marital stability and instability can be organized by the rather simple hypothesis that hostile and hostile–detached couples simply fail to create a stable adaptation to marriage that is either volatile, validating, or avoiding. In other words, the hypothesis is that the longitudinal marital stability results are an artifact of the prior inability of the couple to accommodate to one another and have one of the three types of marriage. For example, in the unstable marriage, a person who is more suited to a volatile or a conflict-avoiding marriage may have married one who wishes a validating marriage. Their influence functions are simply mismatched.

Unfortunately, it is easier to propose this hypothesis than it is to test it. The problem in testing this hypothesis is that the marital interaction is a means for classifying couples. The result of this classification process is that the marriage is described as volatile, validating, or avoiding, rather than describing each person's style or preferences. What is needed to test this hypothesis is an independent method for classifying each person's conflict resolution style. To begin to test this hypothesis, we computed the difference between husbands and wives on the RCISS positive and negative speaker codes. If the mismatch hypothesis was true, one would expect that the results of an analysis of variance between the groups would show greater discrepancies between husbands and wives for the hostile and the hostile–detached groups than for three stable groups. This was indeed the case. By pooling the stable groups into one group and the unstable groups into another group, we obtained the following results: for the positive speaker code, $F(1, 70) = 4.12$, $p < .05$ (stable $= -.01$, unstable $= .08$); for the negative speaker code, $F(1, 70) = 10.42$, $p < .01$ (stable $= -.02$, unstable $= -.26$); and for the difference between positive and negative speaker codes, $F(1, 70) = 8.57$, $p < .01$ (stable $= .01$, unstable $= .34$). Thus, it could be the case that the unstable groups are examples of discrepancies in interactional style between husbands and wives that reflect their differences in preferred type of marital adaptation, or these differences may have emerged over time as a function of dissatisfaction.[9]

This analysis is incomplete without a discussion of the other parameters of our model for these five groups of couples, namely, inertia and influenced and uninfluenced set points. Note that we present no statistical tests here. Our purpose is the qualitative description of the data for generating theory. By *theory* we mean a suggested mechanism for the Gottman–Levenson prediction of marital instability.

Set Points and Inertia

The results in Table 1 summarize the mean set points and inertias for the types of couples. Let us begin by examining the inertia parameter. Nonregulated couples have higher mean emotional inertia than regulated couples; the differences are greater for wives than for husbands (a fourfold difference; .29 vs. .07, respectively). Wives in nonregulated marriages have greater emotional inertia than do husbands, but this is not the case in regulated marriages. Both the influenced and the uninfluenced set points are more negative in nonregulated marriages compared with regulated marriages, and this result is especially true for wives (although we note again that the influenced set point is an attribute of the couple, not the individual). The three stable types of couples also differed from each other. Volatile couples had the highest set points, followed by validators and then avoiders. Also, the effect of influence in nonregulated marriages is to make the set point more negative, whereas, in general, the reverse is true in

[9] Note that this match–mismatch idea is not the same as Watzlawick, Beavin, and Jackson's (1967, p. 67) notion of symmetry versus complimentarity, by which they meant interactional mirroring (as in assertive–assertive) versus one partner complimenting the other (as in assertive–submissive). We would expect, on the basis of the typology reported in Gottman (1994), that couples in which the influence functions are mismatched would differ greatly in their desired levels of emotional distance and closeness and have influence patterns that leave one person feeling overwhelmed and flooded while the other partner feels lonely. Gottman's typology is based extensively on the nature of influence patterns and their concomitants.

Table 1
Parameter Estimates in the Mathematical Modeling of the RCISS Unaccumulated Point Graphs

	Husband set points			Wife set points		
Type	Inertia	Uninfluenced	Influenced	Inertia	Uninfluenced	Influenced
		Regulated couples				
Volatile	.33	.68	.75	.20	.68	.61
Validating	.37	.38	.56	.14	.52	.59
Avoiding	.18	.26	.53	.25	.46	.60
M	.29	.44	.61	.20	.55	.60
		Nonregulated couples				
Hostile	.32	.10	.03	.51	−.64	−.45
Hostile–detached	.40	−.42	−.50	.46	−.24	−.62
M	.36	−.16	−.24	.49	−.44	−.54

Note. RCISS = Rapid Couples Interaction Scoring System.

regulated marriages. Perhaps it is the case that volatile couples need to have a very high set point to offset the fact that they influence one another primarily in the negative range of their interaction. The behavior of the wives was quite different than that of the husbands. Wives in regulated marriages had a set point that was equal to or more positive than husbands'. However, wives in hostile marriages had a set point that was more negative than their husbands', whereas the reverse was true in hostile–detached marriages. The set points of wives in nonregulated marriages were negative and were more negative than the set points of wives in regulated marriages. Wives in hostile marriages had a more negative set point than did wives in hostile–detached marriages.

Validity of the Parameters

The results in Table 2 summarize the correlations of the parameters of our model between the regulated–nonregulated classification, the variables of the dissolution cascade, physical health at Time 2, the summary behavior observation scales of the MICS and the RCISS, and positive affect.

For predicting marital dissolution, these results suggest that (a) the regulated–nonregulated classification (which was the Gottman–Levenson predictor of marital dissolution) was related to the wives' emotional inertia and to both the husbands' and the wives' uninfluenced set points, and (b) the emotional inertia parameters are unrelated to the dissolution cascade variables, but both the husbands' and the wives' uninfluenced set points are related to the dissolution cascade variables. Both the husbands' and the wives' set points are significantly predictive of divorce.

For future research, we would like to know to what extent uninfluenced set points are independent of partner or independent of conversation—that is, to what extent are they intrinsic to the individual and to what extent do they describe a cumulative quality of the relationship?

Emotional inertia. In relation to marital interaction, for the MICS codes, the husbands' inertia was related to their criticism, whereas the wives' inertia was related to the husbands' withdrawal and to the wives' own contempt. For the RCISS codes, the husbands' inertia was related to their contempt, and the wives' inertia was related to all of the subscales of the RCISS.

Set points. For the MICS coding, the husbands' set point variable was related to their criticism, contempt, and withdrawal and to the wives' criticism and withdrawal; for the MICS coding, the wives' set point variable was related to all of the variables for both spouses. For the RCISS coding, the husbands' set point was related to all of their behavior and to all of the wives' behavior except for criticism; the wives' set point was related to all of the husbands' codes except for criticism and to all of the wives' codes.

Positive affect. We observed the following relationships. Wives with more emotional iner-

Table 2
Correlations of Model Parameters With Dissolution Cascade and Behavioral Observation Variables

Variable	Husband inertia	Wife inertia	Husband set point	Wife set point
	Dissolution cascade			
Marital satisfaction (Time 1)				
H	.05	−.08	.25	.26*
W	−.23	−.07	.36*	.28*
Marital satisfaction (Time 2)				
H	−.02	−.21	.31*	.27*
W	.00	−.21	.26*	.20
Thoughts of dissolution				
H considers divorce	.02	.06	−.28*	−.22
H considers separation	.10	.06	−.34**	−.19
H health (Time 2)	−.05	.01	−.31*	−.12
Thoughts of dissolution				
W considers divorce	.00	.24	−.20	−.28*
W considers separation	.02	.24	−.22	−.28*
W health (Time 2)	−.16	.21	.14	−.31*
Dissolution				
Separation	−.07	.01	−.11	−.38**
Divorce	.05	.22	−.32*	−.42***
Marital type	−.04	−.54***	.40***	.54***
	Marital interaction behavior			
MICS Codes				
H criticism	.37**	.21	−.30*	−.27*
H defensive	.11	.07	−.22	−.28*
H contempt	.01	.31*	−.38**	−.31*
H withdrawn	.14	.50**	−.46***	−.61***
W criticism	.11	.19	−.25*	−.27*
W defensive	.09	.21	−.05	−.38**
W contempt	−.01	.36**	−.15	−.30*
W withdrawn	.19	.19	−.30*	−.30*
RCISS Codes				
H criticism	.18	.25*	−.49***	−.24
H defensive	.09	.52***	−.47***	−.46***
H contempt	.26*	.33**	−.74***	−.50***
H withdrawn	−.06	.38**	−.52***	−.55***
W criticism	−.04	.45**	−.23	−.63***
W defensive	.12	.55***	−.30*	−.41***
W contempt	.18	.45***	−.64***	−.63***
W withdrawn	−.08	.31**	−.42***	−.47***
	Positive affect			
H laughter	−.15	−.15	.40***	.13
W laughter	−.08	−.18	.32**	.28*
H smiling	−.01	−.23	.22	.09
W smiling	−.03	−.32*	.26*	.35**

Note. H = husband; W = wife; MICS = Marital Interaction Coding System; RCISS = Rapid Couples Interaction Scoring System.
* $p < .05$. ** $p < .01$. *** $p < .001$.

tia made fewer Duchenne smiles than did wives with less emotional inertia. Husband and wife set points were related to fewer Duchenne smiles, but only for wives. On the RCISS, husbands with higher set points laughed more, whereas wives laughed more when either the husband's or the wife's set point was higher.

Discussion

The purpose of the dynamic mathematical modeling proposed in this article was to generate theory that might explain the ability of the RCISS point graphs to predict the longitudinal course of marriages. We found that the uninfluenced set point, when group averaged, was enough to accomplish this task. This alone is an interesting result. Subsequent attempts at theory construction may profit from making this parameter a function of other dynamic (time varying) variables in the experiment, such as indices of physiological activity. Perhaps the uninfluenced set point represents a cumulative summary of the marriage and reflects what each individual brings to each marital conflict discussion. It might be useful to study what other variables (e.g., stress, coping, and power differences) are related to this index.

Gottman (1994), on the basis of Time 1 interactive behavior on the RCISS, described three distinct types of couples who were more likely to have stable marriages and two groups of couples who were more likely to have unstable marriages. In this article, we examined the influence functions for these five groups of couples and suggested that the influence functions might provide insight into the classification. Validating couples seemed to have a pattern of linear influence over the whole range of their interaction: When they were more negative than positive, they had a negative impact on their partner's subsequent behavior, and, conversely, when they were more positive than negative, they had a positive impact on their partner's behavior. Conflict-avoiding couples, on the other hand, resembled validating couples, but only in the positive ranges of their behavior. In the negative ranges, they had nearly no influence on their spouses. Volatile couples resembled couples headed for marital dissolution in that they had no influence in the positive ranges of their partner's behavior. They differed from this group of couples only in having a more positive uninfluenced set point.

These results provide insight into the potential costs and benefits of each type of stable marriage. The volatile marriage is clearly a high-risk style. Without a high level of positivity, volatile couples may drift to the interactive style of a couple headed for dissolution. The ability to influence one another only in the negative ranges of behavior may suggest a high level of emphasis on change, influence, and conflict in this type of marriage. On the other hand, the conflict-avoiding style seems particularly designed for stability without change and conflict. The validating style seems to combine elements of both styles, with an ability to influence one another across the entire range of interactive behavior. On the other hand, the marriages headed for dissolution had influence functions that were mismatched. In the hostile marriage, the husband, as with a validating husband, influenced his wife in both the positive and the negative ranges but she, as with a conflict avoider, only influenced him by being positive. If we can generalize from validator and avoiding marriages, the wife is likely to seem quite aloof and detached to the husband, whereas he is likely to seem quite negative and excessively conflictual to her. In the hostile–detached marriage, we found another kind of mismatch. The husband, again as with a validating husband, influenced his wife in both the positive and the negative ranges, but she, as with a wife in a volatile marriage, only influenced him by being negative. If we can generalize from validator and volatile marriages, the husband is likely to seem quite aloof and detached to the wife, whereas she is likely to seem quite negative and excessively conflictual to him. These two kinds of mismatches are likely to represent the probable mismatches that might survive courtship. We did not find a volatile style and a conflict-avoiding style within a couple in our data; perhaps they are just too different for the relationship to survive, even temporarily. These results suggest evidence for a mismatch of influence styles in the marriage being predictive of marital instability. This result is interesting in light of the general failure or weak predictability of mismatches in personality or areas of agreement in predicting dissolution (Bentler & Newcomb, 1978; Fowers & Olson, 1986), and it suggests that a study of process may be more profitable in understanding marriage than a study of individual characteristics.

Let us consider what one gains from our approach. As soon as we write down the deterministic model, we already gain a great deal. Instead of empirical curves that predict marital stability or dissolution, we now have a set of concepts that could potentially explain the prediction. We have parameters of uninfluenced set point, influenced set point, emotional inertia, and the influence functions. We have gained a language, one that is precise and mathematical for talking about the point graphs. Marriages that last have more positive uninfluenced set points. Furthermore, interaction usually moves the uninfluenced set points in a more positive direction, except for the case of the volatile marriage, in which the only way anyone influences anyone else is by being negative—in that case a great deal of positivity is needed to offset this type of influence function. Marriages that last have less emotional inertia, they are more flexible and less predictable, and the people in them are more easily moved by their partners. Depending on the type of marriage the couple has, the nature of their influence on one another is given by the shape of the influence functions. We hypothesize that couples headed for divorce have not yet worked out a common influence pattern and that most of their arguments are about differences in how to argue, about differences in how to express emotion, and about differences in issues concerning closeness and distance; all of these differences are entailed by mismatches in influence functions (see Gottman, 1994). Of course, we have no way of knowing from our data whether the mismatches in influence functions were present at the start of the marriage or whether they emerged over time. We are currently studying these processes among newlyweds as they make the transition to parenthood.

As a new methodology for examining an experimental effect and building theory, we submit that the use of these equations is a method that can help a researcher get at the mechanism for an observed effect, as opposed to using a statistical model. A statistical model tells whether variables are related, but it does not propose a mechanism for understanding this relationship. For example, if researchers find that socioeconomic status is related to divorce prediction, they will still have no ideas from this fact how this effect may operate as a mechanism to explain marital dissolution. The differential–difference equation model approach is able to suggest a theoretical and mathematical language for such a theory of mechanism. The mathematical model differs from the statistical model in presenting an equation linking a particular husband and wife over time, instead of a representation of husbands and wives, aggregated across couples as well as time.

The use of the sigmoidal influence function is the next step in developing the model. To accomplish this next step, we need to use an observational system that provides much more data than the RCISS. Gottman (1994) found that the Specific Affect Coding System (SPAFF) is highly correlated with the RCISS speaker slopes, and the advantages of the SPAFF are that the couple's interaction can be coded on-line in real time, without a transcript, and the data are summarized second-by-second instead of at each turn of speech. Thus, the SPAFF will make it possible to obtain much more data for each couple. With the sigmoidal influence function, there is the possibility of five steady states (five intersection points for the null clines), three of which are stable (see the Appendix.) The possible existence of more than one stable steady state for a given couple can be inferred from their data once we have written down the model, which means that we can describe the couple's behavior even in conditions in which they have not been observed in our study. Thus, the model can be used to create simulations of that couple's interaction that go beyond our data.

By varying parameters slightly, we can even make predictions of what will happen to this couple if we could change specific aspects of their interaction, which is a sort of quantitative thought experiment of what is possible for this particular couple. We are currently using this approach in a series of specific intervention experiments designed to change a couple's second interaction about a particular issue. The model can be derived from the couple's first interaction in the laboratory, and the intervention can be designed to change a model parameter (whether it does or not could be assessed). Coupled with an experimental approach, we can test whether the mechanism for change described by the model is accurate by seeing if the model's predictions of what would happen when a model parameter changed is accurate. In this way, the model can be tested and expanded by an interplay of modeling and experimentation.

The qualitative assumptions that form the underpinnings of this effort are also laid bare by the process. For example, the choice of the shape of the influence functions can be modified with considerable effect on the model. Following our qualitative approach, subsequent correlational data can quantitatively test the theory. This can proceed in two ways: (a) the influence functions can be specified in functional (mathematical or graphical) form, and (b) the equations themselves can be made progressively more complex, as needed. To date, our empirical fitting has suggested that the sigmoidal form would best fit the data.

One simple way we suggest changing the equations is to assume that the parameters are not fixed constants but, instead, are functions of other, more fundamental theoretical variables. In the Levenson–Gottman paradigm, there are two central classes of variables we wish to consider. The first class of variables indexes the couple's physiology, and the second class of variables indexes the couple's perception of the interaction derived from our video-recall procedure. We expect that physiological measures that are indicative of diffuse physiological arousal (Gottman, 1990) will be related to less ability to process information, less ability to listen, and greater reliance on behaviors that are more established in the repertoire in upsetting situations (e.g., fight or flight). Hence, it seems reasonable to predict that measures indicative of more diffuse physiological arousal may predict more emotional inertia. Similarly, we expect that a negative perception of the interaction would go along with feeling flooded by the negative affect (see Gottman, 1993) and negative attributions (see Fincham, Bradbury, & Scott, 1990) of one's partner. Hence, it seems reasonable to predict that variables related to the video-recall rating dial would predict the uninfluenced set point. If individuals have an interaction with their spouse that they rate negatively, the next interaction may be characterized by a slightly less positive uninfluenced set point. The uninfluenced set point, to some extent, may index the cumulative effects of the marital balance of positivity over negativity, integrated over time. There is also the possibility that the uninfluenced set point might best be understood by an integration of personality traits with marital interaction patterns.

It is interesting to note that the model is, in some ways, rather grim. Depending on the parameters, the initial conditions determine the eventual slope of the cumulated RCISS curves. Unfortunately, this is essentially true of most of our data. However, another way the model can be developed further is to note that a number of couples began their interaction by starting negatively but then changed the nature of their interaction to a positively sloping cumulative RCISS point graph; their cumulative graph looked somewhat like a check mark. This was quite rare (characterizing only 4% of the sample), but it did characterize about 14% of the couples for at least part of their interaction. This more optimistic type of curve suggests the need to add to the model the possibility of repair of the interaction once it has passed some threshold of negativity. This addition could be incorporated by changing the influence function so that its basic sigmoidal shape had the possibility of a repair jolt (or perhaps "repair nudge" would be closer to the data) in the negative parts of the x-axis of Figure 3. The size of the repair jolts would add two other parameters to the model, each of which would have to be estimated from the data. The jolt would, however, have to be quite sizable to bring the couple far enough away from the zero stable steady state and toward the more positive stable steady state. We might also then inquire as to what the correlates are of these repair jolts. This process would suggest some strength in the marriage that could be explored further.

Finally, the potential precision of the equations suggests experiments in which only one parameter is altered and the effect of the experiment is assessed, thus refining the equation and potentially revealing the structure of the interaction itself. Here is how this would work. After a baseline marital interaction, a standard report based on the observational data would be used to compute the parameters of the model and the influence function. Then, an experiment could be done that changes one variable presumed to be related to the model parameters. For example, we would have participants either relax and lower their heart rates or bicycle until their heart rates increased to 125 beats per minute; then they would have a second interaction, and the model parameters would be recomputed. Order could be counterbalanced. The experiment could reveal the functional relationship between heart rate and the inertia parameter. What is perhaps even more exciting is that the modeling process leads naturally to design experiments. We think that this is so because we are modeling the mechanism. We are building a theory and the theory naturally suggests experiments. Hopefully, the experiments will help build the theory. This process involves both mathematics and the laboratory, which is a new approach in the field of family psychology.

We plan to build this model in subsequent studies by continuous coding that will provide more reliable data for each individual couple, and more of it. These studies will also make it possible to expand from difference to differential equations. The time delay (we used a delay of one time unit in this model) would then become a parameter for each couple; time delays in differential equations are capable of representing cycles. The experiments we are conducting make simulations and subsequent tests of the model possible. What would happen, for example, if we successfully lowered only the couple's heart rate and thus lowered their emotional inertia? Would other parameters of the model change? Would the influence functions change shape? Another development we plan is to study the newlywed couple's transition to parenthood and the effects of the marital conflict on the developing family. When the baby is 3-months-old, we will attempt to model triadic interaction with three equations, perhaps estimating key parameters from the dyadic marital interaction. A system of three nonlinear equations is capable of modeling many complex patterns, including chaos.

References

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

Bentler, P. M., & Newcomb, M. D. (1978). Longitudinal study of marital success and failure. *Journal of Consulting and Clinical Psychology, 46,* 1053–1070.

Ekman, P., & Friesen, W. V. (1978). *Facial Action Coding System.* Palo Alto, CA: Consulting Psychologist Press.

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 Press.

Fowers, B. J., & Olson, D. H. (1986). Predicting marital success with PREPARE: A predictive validity study. *Journal of Marital and Family Therapy, 12,* 403–413.

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

Gottman, J. (1990). How marriages change. In G. R. Patterson (Ed.), *New directions in family research: Depression and aggression.* Hillsdale, NJ: Erlbaum.

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: Erlbaum.

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.

Krokoff, L. J., Gottman, J. M., & Haas, S. D. (1989). Validation of a Rapid Couples Interaction Scoring System. *Behavioral Assessment, 11,* 65–79.

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

Watzlawick, P., Beavin, J. H., & Jackson, D. D. (1967). *Pragmatics of human communication.* New York: Norton.

Weiss, R. L., & Summers, K. J. (1983). Marital Interaction Coding System III. In E. E. Filsinger (Ed.), *Marriage and family assessment.* (pp. 85–115). Beverly Hills, CA: Sage.

Appendix

Linear Stability Analysis

Linear stability analysis is the mathematical technique used to assess the local stability of steady states to small perturbations (see Murray, 1989, for example). Loosely speaking, we determine whether a small "perturbation" away from the steady state will be amplified or damped out over time. A stable steady state will possess a *basin of attraction*, which is defined as those states beginning at which the system would approach the steady state over successive iterations.

We examine the stability of a general steady state, $(\bar{W}, \bar{H})$, to small perturbations. Let us define

$$W_t = \bar{W} + w_t \qquad H_t = \bar{H} + h_t \tag{A1}$$

where w_t and h_t are small.

Then, using Equation 3 and Equation 4, expanding in a Taylor series and retaining only linear terms, we have

$$W_{t+1} = \bar{W} + w_{t+1} = I_{HW}(\bar{H} + h_t) + r_1(\bar{W} + w_t) + a \approx I_{HW}(\bar{H}) + h_t \frac{d}{dH} I_{HW}(\bar{H}) + r_1(\bar{W} + w_t) + a. \tag{A2}$$

But $I_{HW}(\bar{H}) + a = (1 - r_1)\bar{W}$ by definition of $\bar{W}$. Hence, this equation simplifies to

$$w_{t+1} = r_1 w_t + h_t \frac{d}{dH} I_{HW}(\bar{H}). \tag{A3}$$

Similarly, $h_{t+1} = r_2 h_t + w_{t+1}\, d/dW\, (I_{WH}[\bar{W}])$, which on substitution for w_{t+1} leads to the following pair of linear difference equations:

$$\begin{bmatrix} w_{t+1} \\ h_{t+1} \end{bmatrix} = \begin{bmatrix} r_1 & \dfrac{dI_{HW}}{dH} \\ r_1 \dfrac{dI_{WH}}{dW} & r_2 + \left(\dfrac{dI_{WH}}{dW}\right)\left(\dfrac{dI_{HW}}{dH}\right) \end{bmatrix} \begin{bmatrix} w_t \\ h_t \end{bmatrix}. \tag{A4}$$

The stability matrix is therefore

$$M = \begin{bmatrix} r_1 - 1 & \dfrac{dI_{HW}}{dH} \\ r_1 \dfrac{dI_{WH}}{dW} & r_2 + \left(\dfrac{dI_{WH}}{dW}\right)\left(\dfrac{dI_{HW}}{dH}\right) - 1 \end{bmatrix}, \tag{A5}$$

with stability requiring $\mathrm{Tr}(M) < 0$ and $\mathrm{Det}(M) > 0$. These conditions for linear stability reduce to the following:

$$r_1 + r_2 + \left[\frac{d}{dW} I_{WH}(\bar{W})\right]\left[\frac{d}{dH} I_{HW}(\bar{H})\right] < \min(2, 1 + r_1 r_2) \tag{A6}$$

or

$$\left[\frac{d}{dW} I_{WH}(\bar{W})\right]\left[\frac{d}{dH} I_{HW}(\bar{H})\right] < \min[(1 - r_1)(1 - r_2), (1 - r_1) + (1 - r_2)]. \tag{A7}$$

If the inertia values, r_1 and r_2, are less than one (the observed values generally are), this equation reduces to

$$\left[\frac{d}{dW} I_{WH}(\bar{W})\right]\left[\frac{d}{dH} I_{HW}(\bar{H})\right] < (1 - r_1)(1 - r_2). \tag{A8}$$

Thus, to assess stability, we must simply evaluate the derivatives of the influence functions and use Equation A7. In general, we can say that steep influence functions and high inertia (r_i) are destabilizing. For example, if each influence function had a slope greater than one, then the steady state would be unstable irrespective of the inertia values. This hypothesis agrees with our intuitive expectations if we interpret instability as the amplification of small perturbations: Influence is a measure of the effect one partner has on another, so that large (positive) changes in influence will result in mutual amplification or instability. On the other hand, even couples with relatively flat (low derivative) influence functions can have unstable steady states if either of the partners' inertia is high (i.e., close to one).

We now examine the special case in which the inertias, r_1 and r_2, are less than one but greater than zero. Then, $(1 - r_1)(1 - r_2) < (1 - r_1) + (1 - r_2)$, so that a steady state is stable if and only if

$$\left[\frac{d}{dW} I_{WH}(\bar{W})\right]\left[\frac{d}{dH} I_{HW}(\bar{H})\right] < (1 - r_1)(1 - r_2). \tag{A9}$$

This equation can be interpreted graphically. The null clines will intersect either as shown in Figure A1a or A1b. From Equations 3 and 4, the equations for the

(*Appendix continues on next page*)

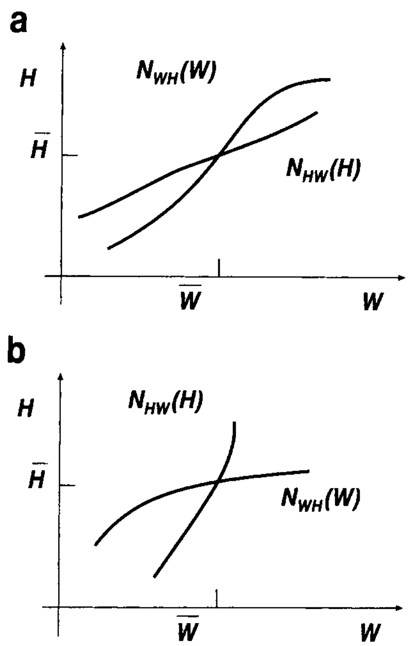

Figure A1. Null clines and the stability of steady states. Steady states correspond to points at which the null clines intersect. The stability of a steady state can be determined graphically (see text for details): When the null clines intersect as is shown in a, the steady state is unstable; when they intersect as is shown in b, the steady state is stable. H = Husband; W = Wife.

null clines can be written

$$N_{HW}(H) = [I_{HW}(H) + a]/(1 - r_1)$$
$$N_{WH}(W) = [I_{WH}(W) + b]/(1 - r_2). \quad \text{(A10)}$$

The derivative of N_{HW} is $d/dH\ I_{HW}(H)/(1 - r_1)$, so that the tangent at $(\bar{W},\bar{H})$, when plotted as in Figure A1a and A1b, has slope $(1 - r_1)/d/dH\ I_{HW}(\bar{H})$. Similarly, the tangent to the N_{WH} null cline has slope $d/dW I_{WH}(\bar{W})/(1 - r_2)$.

If the null clines intersect as shown in Figure A1a, then clearly

$$(1 - r_1)/\frac{d}{dH} I_{HW}(\bar{H}) < \frac{d}{dW} I_{WH}(\bar{W})/(1 - r_2). \quad \text{(A11)}$$

Together with the assumption concerning the inertia, this implies

$$\frac{d}{dH} I_{HW}(\bar{H}) \frac{d}{dW} I_{WH}(\bar{W}) > (1 - r_1)(1 - r_1), \quad \text{(A12)}$$

and hence by Equation A9 the steady state is unstable.

If, on the other hand, the null clines intersect as in Figure A1b, then the inequality is reversed and the steady state is stable. Thus, under these conditions on the inertia, we can determine graphically not only the location of steady states but also their stability.

A simple corollary of this null cline intersection rule is that stable and unstable steady states must alternate—that is, any two stable steady states are separated by an unstable one and vice versa. If it is assumed that influence functions are monotonic increasing functions, then the steady states can be ordered. By this we mean that the steady state values, $\bar{W}$ and $\bar{H}$, will both increase as we move from one steady state to the next. If we assume that influence functions saturate, then highest and lowest steady state are clearly stable (they must intersect as in Figure A1b). We can thus categorize the possible steady states and their stability: There must be an odd number of steady states that alternate between stable and unstable, and the first and last are stable.

Received December 14, 1993
Revision received October 20, 1994
Accepted November 20, 1994 ■