3

Threshold Models of Diversity: Chinese Restaurants, Residential Segregation, and the Spiral of Silence

Mark Granovetter and Roland Soong†*

In many binary decisions, a person's choice depends in part on the composition of the group that has already made one or the other choice. In deciding whether to live in a neighborhood, a person may consider the ethnic composition of the neighborhood. In deciding whether to speak out on a public issue, a person may consider the proportion of previously expressed opinions that are the same as his. Substantial literatures have grown up around these two examples, which go under the rubrics of residential tipping and pluralistic ignorance. We develop a mathematical model that applies to all such binary situations and illustrate it especially by the examples of residential segregation and public opinion. The model builds on and generalizes previous work on these subjects, and it is related to but distinct from the authors' earlier work on threshold models of collective behavior. We conclude with a report on preliminary attempts at empirical measurement.

The authors are listed alphabetically. Early drafts of the paper were prepared by the first author in sabbatical facilities provided by the Harvard University Department of Sociology, and a final revision has benefitted from facilities and services provided by the Stanford University Graduate School of Business during the first author's visiting appointment. Helpful comments from D. Garth Taylor, Steve Rytina, Thomas Schelling, and Harrison White have improved the paper. Partial support of the work was provided by a John Simon Guggenheim Memorial Foundation Fellowship and by NSF grant SPI 81-65055 to the first author.

*State University of New York at Stony Brook
†Arbitron Ratings Company

1. INTRODUCTION

How you choose between two alternatives may depend in part on how others have chosen before you. This dependence may involve the proportion or number (in some reference group) who have previously made one or the other decision. Whether you join a riot or buy a product may hinge on the *proportion* of others who have already done so; then, how many have *not* rioted or bought is as important as how many have. In some situations, absolute numbers seem more important. Whether you turn on your headlights at some given level of daylight probably depends on the *number* of other drivers you have seen do so; those who have not may be ignored. This dependence on either proportions or numbers can be modeled with what we have called threshold models of collective behavior (Granovetter 1978; Granovetter and Soong 1983, 1986).

There are situations typical of socially differentiated populations that require a different model. As with absolute numbers, the focus may only be on that part of the reference group that has made a particular decision, but there may be a special concern with the *group composition* of that part. Few would notice what kind of person turns on his headlights, but many might avoid a riot involving whites and blacks until their own group exceeded a certain proportion of the rioters. Would-be customers of an ethnic restaurant may be very attentive to the proportion of the restaurant's clientele that is from the relevant ethnic group. Here, one attends only to the composition of those who are eating, partly because the set of those who chose not to eat in the restaurant is not easily visible, though it may be well defined. Or consider a club whose members are drawn from some definite, ethnically heterogeneous, eligible population. Individuals may be willing to join only when the proportion of current members who are from their *own* group exceeds some minimum; they pay no attention to those eligible but outside the club.

These situations arise when the social composition of those making a decision—to enter a riot, restaurant, or club—is considered important, for whatever reasons. In clubs, pure social snobbery may be rationalized by a belief that only members of one's own group are congenial companions. This belief is usually grounded in real or imagined barriers of language, culture, or behavior. In an ethnic restaurant, the proportion of diners from the relevant ethnic group may be taken as a signal (Spence 1974) of the quality and authenticity of the cuisine. In deciding whether to engage in some task, individuals may be influenced by a belief that technical competence is available in only one of the groups, or that one can best function cooperatively, keep secrets, and act unselfishly in dangerous situations with co-ethnics. Thus, units of spies and soldiers have often been ethnically homogeneous. However, purely instrumental calculations can also lead to preferring that a group other than one's own be dominant, as when minority parents attempt to enroll their children in a school with a large proportion of the majority group, suspecting that it will be favorably endowed with resources.[1]

2. RESIDENTIAL TIPPING AND THE SPIRAL OF SILENCE

Our interest here is not in explaining such preferences but in constructing a formal model that illuminates their consequences. As we do so, we will refer repeatedly to two illustrations, each with a substantial associated literature. They are substantively quite different from one another and have not previously been discussed together. The differences will be useful in indicating the range of phenomena that can be subsumed under this model and in suggesting some of the modifications needed for particular applications.

The first illustration is residential segregation. Thomas Schelling's seminal papers (1971, 1973, 1978) develop a model for bounded neighborhoods. Each individual resides in the neighborhood "unless the percentage of residents of opposite color exceeds some limit. Each person, black or white, has his own limit. ...If a person's limit is exceeded in this area he will go somewhere else—a place, presumably where his own color predominates or where color does not matter" (1971, p. 167). We use Schelling's account to help generate our formal

[1] When interest centers on the proportion from one particular group, it is irrelevant whether that group is one's own, since the population can then be divided into two categories: the particular group and all the others. Joining the riot only when r percent of the rioters belong to one's own group is exactly equivalent to joining only when $(1 - r)$ percent are from the other group(s). For simpler exposition, we will usually speak of two groups; we will show that generalization to n groups is conceptually (if not computationally) straightforward.

model, which subsumes his results and permits important generalizations.

The second illustration comes from the literature on public opinion. Discussions of "pluralistic ignorance" have observed that individuals may fail to speak out on important issues because they falsely perceive their own opinion to be in the minority. Thus, "moral principles with relatively little popular support may exert considerable influence because they are mistakenly thought to represent the views of the majority, while normative imperatives actually favored by the majority may carry less weight because they are erroneously attributed to a minority" (O'Gorman and Garry 1976, p. 450). Noelle-Neumann (1974, 1977, 1984) has introduced some dynamics to this literature with the theory of a "spiral of silence." Her model was inspired by a situation in West Germany in the late 1960s, when followers of the Christian Democratic Party didn't express themselves publicly but Socialist supporters did. This "encouraged people either to proclaim their views or to swallow them and keep quiet until, in a spiraling process, the one view dominated the public scene and the other disappeared from public awareness as its adherents became mute" (1984, p. 5). Taylor suggests that "one's perception of the distribution of public opinion motivates one's willingness to express political opinions. The act of self-expression, however, changes the global environment of opinion, altering the perceptions of other persons and, ultimately, affecting their willingness to express their own opinions" (Taylor 1982, p. 311, 1986).

Noelle-Neumann's extensive public opinion research in Germany has confirmed in many different ways that individuals have definite views on what others believe and that their assessments of those beliefs affect their own willingness to speak out (1984, chap. 2). Similarly, in an analysis of survey data on conflict in Boston over court-ordered busing to achieve school desegregation, Taylor (1986) notes that the extent to which people opposed to busing expressed this opposition—in ways ranging from discussing it with their friends and neighbors to supporting illegal protest actions and boycotts—depended strongly on the extent of opposition to busing that they *perceived* among their neighbors. For example, 56 percent of those who thought none of their neighbors agreed with the court decision discussed their (negative) opinions on busing frequently or very frequently with them, 41 percent of those who thought 10 percent or so of their neighbors agreed with it did so, and 28 percent of those who thought 25 percent or more of their neighbors agreed with it did so.

Our formal model incorporates many of these insights but drops the assumption—prominent especially in Noelle-Neumann's work—that concern about being in the majority or minority has a special status. Instead, we offer a more general argument that each individual has some sensitivity to the predominance of his own opinion among those previously expressed but that these sensitivities may vary continuously. As in previous work on threshold models, small changes in the distribution of sensitivity will have large impacts on equilibrium outcomes. This may help explain why "the intensity of anti-busing protest varies even though the level of opposition to busing is relatively constant in American cities" (Taylor 1983, p. 21).

Both cases illustrate the situation of interest to us. In the first, people are sensitive to the racial proportions in a neighborhood and reside there if those proportions are suitable. In the second, people express their opinion if it agrees with that of some proportion of those who have previously expressed themselves. The racial makeup of those interested in the neighborhood but not living in it and the opinions of those who are silent do not count, partly because they cannot easily be determined. A formal model should predict the exact composition of the neighborhood, or of expressed public opinion, and whether that composition settles down to some equilibrium. In all the examples we have given, the size, homogeneity, and diversity of the final outcome are of special interest. Therefore, for simplicity, we refer to our models as models of diversity.

3. THRESHOLD MODELS OF DIVERSITY

Suppose that each individual belongs to one of two groups and is characterized by a threshold, i.e., the proportion he would have to see of all those choosing one side of a binary decision who are in his own group before he would also make that same choice. For example, a black man is said to have a threshold of 35 percent if he is unwilling to live in a neighborhood unless it is at least 35 percent black. An opponent of nuclear power who would not speak up until 60 percent of expressed public opinion agreed with his view has a threshold of 60 percent. (Note that *threshold*, as it is used here, differs from its use in Granovetter [1978] and in subsequent work on threshold models, because those making one side of the binary decision are ignored. Strictly speaking, we should use some qualifier like *diversity-type thresholds*, but we hope the difference will be clear from context.)

Whites and blacks cannot change their color, but individuals can change their opinions. The model we develop here abstracts from such change and is thus best suited to the study of cases in which expressions of opinion may change but the actual opinions do not. Such stability of opinions may be common. In his book on the Boston school busing controversy, Taylor (1986) notes that the proportion opposed to busing was quite stable over 22 months and five waves of surveys, beginning at 86 percent in wave 1 and varying from 89 to 91 percent from waves 2 to 5. The rancorous conflict of the period did not much affect opinions. Moreover, though individuals cannot change their color, the population racial composition does change, and this may have a similar impact on the workings of the model.

Our notation and discussion will stress the segregation interpretation, since it would be awkward and redundant to develop two separate stories for each equation and analysis. The reader with special interest in public opinion should be able to supply the relevant story in each case, and we will draw on this interpretation for contrast and generality. Assume (as Schelling does) that there is a fixed number of whites, N_w, and blacks, N_b, available to live in the neighborhood and that any number may live there in a given time period. Then, neighborhood population may fluctuate drastically over several periods and may at times be reduced to zero. This fluctuation is one reason that the neighborhood interpretation is not entirely natural, though we will introduce capacity constraints later to make it more plausible. Such fluctuations are easier to imagine for the public opinion interpretation, in which the transaction costs of entering or leaving the set of those expressing their opinions are low.

Outcomes are determined by the exact distributions of thresholds. We will refer especially to the cumulative distribution functions (cdf's) for black and white thresholds. Thus, $F_w(p_w)$ gives the proportion of whites whose threshold is less than or equal to the proportion of whites in the neighborhood, given by p_w. For example, if $F_w(0.25) = 0.65$, 65 percent of the whites have thresholds less than or equal to 25 percent. Then, if the neighborhood is at some time exactly 25 percent white, 65 percent of the white population (including those currently nonresident) will be willing to live in it. If this is more than the number of whites already in the neighborhood, there will be an influx of whites; if less, there will be an outflux. For blacks, the corresponding cdf is $F_b(p_b)$.

We set up the dynamics in discrete time: At the beginning of each period, each individual, black or white, in or out of the neighborhood, observes the racial makeup of the neighborhood in the previous period. If his threshold is met or exceeded (i.e., if the proportion of his own group in the neighborhood is equal to or higher than his minimum proportion), he will reside in the neighborhood that period; otherwise, he will not.[2] This leads us to a pair of coupled first-order difference equations. Let $n_w(t)$ be the number of whites in the neighborhood at time t, and let $n_b(t)$ be the number of blacks. The proportion of whites at t is then $p_w(t) = n_w(t)/[n_w(t) + n_b(t)]$. Suppose the neighborhood is 25 percent white at time t. What proportion of the neighborhood will be white at $t+1$? Those whites resident at $t+1$ will be exactly those whose thresholds are less than or equal to 0.25. If $F_w(0.25)$ were 0.65, and if N_w were the total population of whites available to live in the neighborhood, then at $t+1$ we would have $0.65N_w$ whites in residence. This is the same as saying that

$$n_w(t+1) = N_w F_w[p_w(t)]. \tag{1a}$$

By the same reasoning, we have for blacks

$$n_b(t+1) = N_b F_b[p_b(t)]. \tag{1b}$$

In general, we expect the cdf's to be nonlinear, and exact solution of the system for explicit time paths will rarely be possible, though system equilibria may nevertheless be found explicitly. By forward recursion we can always trace out any desired segment of the time path. Equilibrium requires that $n_w(t+1) = n_w(t)$ and that $n_b(t+1) = n_b(t)$. Call these equilibrium numbers of whites and blacks w and b. Substituting into equations (1a) and (1b) gives us as conditions for equi-

[2] Those who shift in or out of the neighborhood during this period do so in no particular order. We do not make Schelling's "somewhat plausible assumption that, as between two whites dissatisfied with the ratio of white to black, the more dissatisfied leaves first—the one with lesser tolerance" (1971, p. 168). Schelling asserts that this assumption is needed for his results, but our model will reproduce them in detail without it. A weaker version of the assumption follows from our setup: If in two successive periods there is a net outflow of whites, all those leaving in the first period are more tolerant (i.e., have a lower threshold) than all those leaving in the second.

librium

$$w = N_w F_w [w/(w+b)], \tag{2a}$$

$$b = N_b F_b [b/(w+b)]. \tag{2b}$$

Any admissible pairs (w, b) that satisfy these equations are equilibria. These can always be approximated by simultaneously graphing the two equations and noting all intersections.

Whether a particular equilibrium actually occurs depends on its stability. An equilibrium is asymptotically stable if the system moves back toward it after a slight displacement; it is unstable if the system moves further away. Thus, stable equilibria attract and unstable equilibria repel all nearby trajectories. We expect, in practice, to see only stable equilibria. We may assess the stability of an equilibrium by linearization, i.e., by taking the system as linear near the equilibrium point and approximating its behavior by a Taylor expansion that omits terms of order 2 or higher. This yields the following test for stability (see, e.g., Luenberger 1979, pp. 324–28). Consider the matrix of first partial derivatives of the system, evaluated at an equilibrium. That equilibrium is stable if and only if all eigenvalues of the matrix are strictly less than unity and unstable if any is greater. The test fails (is uninformative) if none exceeds unity but one or more are equal to it. Define w_0 as $n_w(t)$, w_1 as $n_w(t+1)$, b_0 as $n_b(t)$, and b_1 as $n_b(t+1)$. Then, the relevant matrix of partials is

$$\begin{bmatrix} \partial w_1/\partial w_0 & \partial w_1/\partial b_0 \\ \partial b_1/\partial w_0 & \partial b_1/\partial b_0 \end{bmatrix}.$$

Recall that w and b denote the equilibrium numbers of blacks and whites, and let f_w be the probability density of thresholds for whites and f_b the probability density for blacks. (These are just the usual first derivatives of the cdf's F_w and F_b.) Computing the partial derivatives above and evaluating them at some equilibrium point (w, b) then yields

$$\begin{bmatrix} N_w f_w[w/(w+b)]\left[b/(w+b)^2\right] & -N_w f_w[w/(w+b)]\left[w/(w+b)^2\right] \\ -N_b f_b[b/(w+b)]\left[b/(w+b)^2\right] & N_b f_b[b/(w+b)]\left[w/(w+b)^2\right] \end{bmatrix}.$$

One eigenvalue of this matrix is zero; stability then depends on whether the nonzero eigenvalue is less than unity. This eigenvalue is

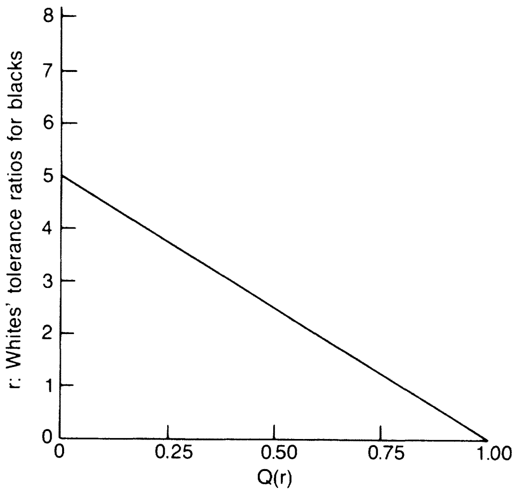
$$N_w f_w[w/(w+b)]\left[b/(w+b)^2\right] + N_b f_b[b/(w+b)]\left[w/(w+b)^2\right]. \tag{3}$$

It is qualitatively clear from this expression that the smaller the absolute sizes of the two groups and the larger the numbers of individuals in the neighborhood at equilibrium, the greater the likelihood of stability. For the public opinion interpretation, this implies that the smaller the number holding an opinion and the larger the number expressing one, the greater the likelihood of a stable distribution of expressed opinion. It follows that, other things equal, an issue on which large numbers have an opinion but few express it will engender volatile distributions of public expression. More stability will occur in a small population in which many express their views.

4. USING THE MODEL: STRAIGHT-LINE TOLERANCE DISTRIBUTIONS

We first apply the model to the straight-line tolerance distributions that are Schelling's main illustrations (1971, pp. 167–75; 1978, pp. 155–64), since these have received wide attention. Our more formal account clarifies the interplay between model parameters and neighborhood outcomes, allowing assessment of the parameter ranges over which one can expect Schelling's counterintuitive results. Our Figure 1 is comparable to the first part of Schelling's Figure 18 (1971, p. 169; or see Schelling 1978, fig. 9, p. 158), except that the y intercept

FIGURE 1. $Q(r)$ = proportion of whites with tolerance ratios greater than or equal to r.



is at 5 rather than 2 (see Schelling 1971, p. 171, or 1978, p. 161). On our y axis are what Schelling calls the tolerance ratios of whites for blacks, i.e., the maximum ratio of blacks to whites that whites can accept and still live in the area. Like Schelling, we assume that an identical distribution governs blacks' tolerance ratios for whites. On the x axis is the proportion of whites with ratios greater than or equal to that given by the y coordinate. Thus, the point (0.25, 3.75) indicates that 25 percent of whites have a black-to-white tolerance ratio of 3.75 to 1, or greater.[3] Denote the proportion of individuals with tolerance ratios greater than or equal to some given ratio r by $Q(r)$. Generalizing the straight-line distribution of Figure 1, suppose the line intersects the y axis at some value, R. By definition, the line must intersect the x axis at unity, so its slope is just $-R$. The equation for the line is then $r=(-R)Q(r)+R$, so $Q(r)=1-(r/R)$.

A white who would live in the area as long as the proportion of blacks was less than or equal to one in four has a tolerance ratio of 1/3; but this is, in our terms, a threshold of 0.75. If some proportion of the population have tolerance ratios greater than or equal to a given r, then this proportion have thresholds less than or equal to the corresponding s. For example, those with tolerance ratios greater than or equal to 1/3 have thresholds less than or equal to 0.75. They will not live in a neighborhood unless three out of four of its residents are their own color. Denote the cdf of thresholds by F, the thresholds themselves by s, and the tolerance ratios by r. Then, $r=(1/s)-1$.[4] It follows that in general, $F(s)=Q(r)=Q[(1/s)-1]$. Substituting in our equation for the straight-line tolerance distribution, we have $F(s) = 1-\{[(1/s)-1]/R\}=(1/R)(1+R-(1/s))$ as the cdf of thresholds for a straight-line tolerance distribution that intersects the y axis at R. Note, however, that for $r>R$, $Q(r)=0$: By hypothesis, no one has a

[3] Schelling refers to this curve as a cumulative distribution of tolerances, but this is confusing, since cumulative distributions usually indicate proportions less than or equal to the value of the random variable. It may also confuse readers that the axes are reversed from the usual depiction of cdf's, perhaps a carryover of the axis reversal practiced by economists in supply-and-demand schedules, in which price, the independent variable, is on the y axis.

[4] We prefer thresholds to tolerance ratios because ratios make it awkward to discuss those willing to live in a neighborhood or express an opinion before other members of their group have done so; i.e., the ratio requires division by zero. In our terms, such individuals simply have a threshold of zero. Thresholds are identical to what Schelling (1973) called individual tipping points.

tolerance ratio greater than R. Correspondingly, for $s < (1/(1+R))$, we have $F(s) = 0$.

Thus, F is defined piecewise over the argument s, and though straight-line tolerance distributions appear simple at first, the piecewise property implies that the pair of difference equations operating will vary depending on the exact values of the variables at a given time. For example, from equation (1a), we have the relation $n_w(t+1) = N_w F_w(p_w(t))$. But to evaluate this relation here, we must note that when the argument of F is less than $(1/(1+R))$, $F = 0$. In this case, $p_w(t) < (1/(1+R))$, which occurs if and only if $[n_b(t)/n_w(t)] > R$. When F_w is zero, the difference equation for whites reduces to $n_w(t+1) = 0$. As the distribution of tolerances implies, when the current ratio of blacks to whites exceeds R, no white will want to live in the neighborhood. When the argument of F is greater than or equal to $(1/(1+R))$, we have $F_w(p_w(t)) = (1/R)[1 + R - (1/p_w(t))] = 1 - [n_b(t)/Rn_w(t)]$ and a corresponding equation of $n_w(t+1) = N_w\{1 - [n_b(t)/Rn_w(t)]\}$. The analysis is exactly symmetrical for blacks. It follows that there are four different ways to define the system of difference equations:

1. If $n_b(t)/n_w(t) > R$, $n_w(t)/n_b(t) \le R$, then

$$n_w(t+1) = 0, \tag{4a}$$

$$n_b(t+1) = N_b\{1 - [n_w(t)/(Rn_b(t))]\}. \tag{4b}$$

2. If $n_b(t)/n_w(t) \le R$, $n_w(t)/n_b(t) > R$, then

$$n_w(t+1) = N_w\{1 - [(n_b(t)/Rn_w(t))]\}, \tag{4c}$$

$$n_b(t+1) = 0. \tag{4d}$$

3. If $n_b(t)/n_w(t) > R$, $n_w(t)/n_b(t) > R$, then

$$n_w(t+1) = 0, \tag{4e}$$

$$n_b(t+1) = 0. \tag{4f}$$

4. If $n_b(t)/n_w(t) \le R$, $n_w(t)/n_b(t) \le R$, then

$$n_w(t+1) = N_w\{1 - [n_b(t)/(Rn_w(t))]\}, \tag{4g}$$

$$n_b(t+1) = N_b(1 - \{n_w(t)/[(Rn_b(t))]\}. \tag{4h}$$

For each situation, note that at equilibrium, $n_w(t+1) = n_w(t) = w$ and $n_b(t+1) = n_b(t) = b$. In situation 1, w must $= 0$; substituting into (4b) gives $b = N_b$. It is easily verified that $(0, N_b)$ satisfies (4a) and (4b). The eigenvalue from equation (3) is then $F_w(w/(w+b)) = F_w(0) = 0$, so $f_w(0)$ also $= 0$. This zero eigenvalue indicates stability.

Situation 2 is exactly symmetrical to situation 1, so we have a second stable equilibrium $(N_w, 0)$. In situation 3, each group finds too many of the other, and everyone leaves the neighborhood. The equilibrium $(0,0)$ satisfies the two equations, but derivatives of any order are equal to zero, and linearization fails in this degenerate case: No Taylor approximation provides any information unless the relevant matrices of derivatives are nonzero. But we can see by inspection the stability characteristics of this equilibrium. If a perturbation results in numbers of whites and blacks within the parameters of situation 3, the equilibrium is stable, because the system then returns to $(0,0)$. If a perturbation results in a situation other than 3, the system moves away from that equilibrium.[5] But stable equilibria in other situations must also be defined in this way: Any perturbation that takes the system out of the parameter region within which an equilibrium was defined leads the system to a stable equilibrium within the new region. Therefore, we can simply say that the equilibrium of region 3 is stable.

In situation 4, the system is governed by a pair of nontrivial equations. Substituting the equilibrium conditions into (4g) and (4h) gives

$$w^2 = (N_w/R)(Rw - b), \tag{4i}$$

$$b^2 = (N_b/R)(Rb - w). \tag{4j}$$

We can solve (4i) for b, yielding $(Rw) - (R/N_w)w^2$, and substitute this

[5] But the system will not enter situation 3 unless $R < 1$. This can be shown as follows: Suppose at time t we have $n_w(t)$ whites and $n_b(t)$ blacks. If $n_w(t) = n_b(t)$, then situation 3 can occur only if $R < 1$. If $n_w(t)$ does not $= n_b(t)$, suppose that $n_w(t) > n_b(t)$. Then, $n_w(t)/n_b(t) > 1$ and $n_b(t)/n_w(t) < 1$; thus, they can both be greater than R only if $R < 1$. It follows immediately that if R is less than 1, both ratios cannot be less than 1, so that in this circumstance, region 4 cannot be entered. Thus, for any given value of R, the system may enter either regions 1, 2, and 3 ($R < 1$) or regions 1, 2, and 4 ($R \geq 1$). If we plot $n_b(t)$ on the vertical axis and $n_w(t)$ on the horizontal when $R < 1$, the region of stability for the equilibrium in region 3 $(0,0)$ is bounded by the two rays from the origins given by $n_b(t) = n_w(t)/R$ on the left and by $n_b(t) = Rn_w(t)$ on the right. To the left of the first ray, we are in region 1; to the right of the second ray, we are in region 2. As R approaches 0, the middle, stable region for the equilibrium of $(0,0)$ increases, and the size of the other two regions approaches zero asymptotically. This is intuitively reasonable, since a very small value of R implies that each group finds the other nearly intolerable and everyone leaves in droves, emptying the neighborhood.

into (4j). This produces a fourth-degree equation for w:

$$0 = w\left\{\left(R^2/N_w^2\right)w^3 - \left(2R^2/N_w\right)w^2 + \left[R^2 + (N_b/N_w)R\right]w + \left[(N_b/R) - N_b R\right]\right\}. \tag{4k}$$

Since the situation for blacks is symmetrical, the same equation (with interchanged subscripts) applies to b. Each of these equations has four roots. For a corresponding pair (w, b) to be an admissible equilibrium, it must fall within the scope of situation 4, w and b must both be real and positive, and we must have $w < N_w$ and $b < N_b$. One root $(0,0)$ is outside the boundaries of situation 4, so we have a maximum of three equilibria here.

Some formidable algebra would yield a general expression for the three equilibria, but it is simpler to use approximations for specified parameters. One can either plot equation (4k) and note the approximate location of zeroes or go back to equations (4i) and (4j) and plot them together, noting the intersections.[6] In the special case in which $R = 5$ and $N_w = N_b = 100$ (as in Schelling 1971, p. 171, fig. 19; 1978, pp. 161–62, fig. 10), the three solutions are found to be approximately (25.36, 94.64), (80, 80), and (94.64, 25.36). All are valid equilibria and can be seen to work by substitution back into the original difference equations. The eigenvalue test shows that the first and third are asymptotically unstable and the middle is stable. For this set of parameters, then, we have three stable equilibria (one each from situations 1, 2, and 4) and two unstable equilibria (from situation 4). The $(0,0)$ equilibrium of region 3 is irrelevant here, because when $R > 1$, this region cannot be entered.

5. A GRAPHICAL METHOD

Some of these results can be seen graphically by generalizing a method suggested by Schelling (1971, p. 169). He draws curves that depict the maximum number of the other group that any fixed number of a group can tolerate. To determine the maximum number of whites

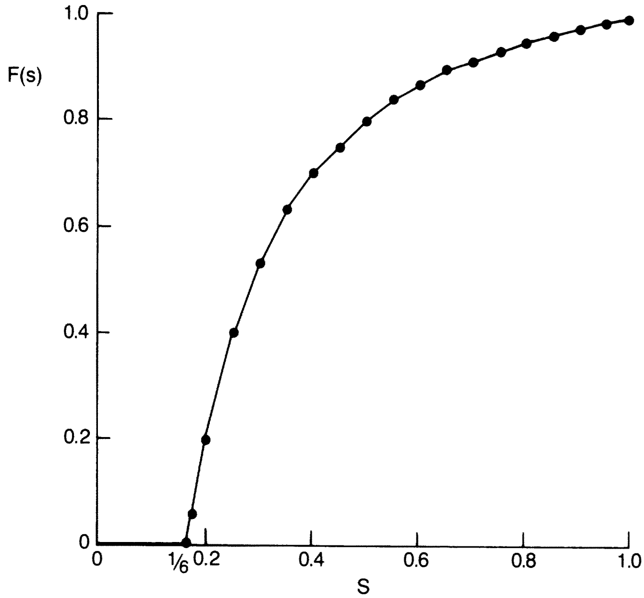
[6] This is, in effect, what Schelling has done in Figures 18–29 (1971) and Figures 9–12 (1978, pp. 158–64), though he does not interpret his procedure as simultaneous graphing of equilibrium equations and does not note those intersections that correspond to unstable equilibria. See our further discussion in the next section on graphical methods.

that a given number of blacks could tolerate (i.e., would be willing to live in a neighborhood with), find the black with the *highest* threshold, since that individual has the least tolerance for whites. (We assume that the b blacks in question consist of the black with the lowest threshold, the black with the next lowest, and so on, up to the bth from the bottom of the distribution. This follows from the definition of the task, which is to determine the maximum number of whites that exactly b blacks could live with.) For the public opinion interpretation, the curves indicate the largest number of those with opposite opinions that some fixed number could tolerate and still express their opinion.

By definition of $F(s)$, the cdf of thresholds, the highest threshold s, represented among k percent of the population, is the s that satisfies the equation $F(s) = k$. Here, we have $k = b/N_b$ and $s = [b/(w + b)]$. But when b, N_b, and F are already given, solving this equation yields not only the highest threshold among blacks, but also w, precisely the maximum number of whites the b blacks could tolerate. Suppose, for example, that we wanted to know how many whites the most tolerant 70 in a population of 100 blacks could live with and that the cdf indicated that $F_b(0.40) = 0.70$, i.e., that 70 percent of the black population is willing to live in a neighborhood that is 40 percent or less black or, similarly, that the highest threshold represented within that part of the black population is 40 percent. (This example is drawn from a straight-line tolerance distribution, with $R = 5$. See the cdf in Figure 2.) Thus, all these 70 blacks will be satisfied as long as $[70/(70 + w)] \geq 0.40$, i.e., when $w \leq 105$. The graphical procedure for constructing the curve is as follows. For a given number b, find the proportion (b/N_b) on the vertical axis of the cdf, draw a horizontal line to the cdf, and drop a line to the x axis to find the argument of F that corresponds to that proportion. Once that argument is found, the maximum number of whites can be computed. For each b, then, there is some number w, a function of b that we will call $e(b)$. Here, $105 = e(70)$.

For the 70 blacks and 105 whites in this example, blacks are just 40 percent of the neighborhood, so all blacks are satisfied. But no more blacks will enter the neighborhood, since the next black will, by definition of cdf, have a threshold higher than 40 percent. For blacks, then, the points on $e(b)$ represent an equilibrium situation—thus the notation "e." Therefore, it is not surprising that the equation $b/N_b = F[b/(w + b)]$, from which we derived points for $e(b)$, is exactly the

FIGURE 2. $F(s) = (1/R)[1 + R - (1/s)]$, where $R = 5$: The cdf of a straight-line tolerance distribution.

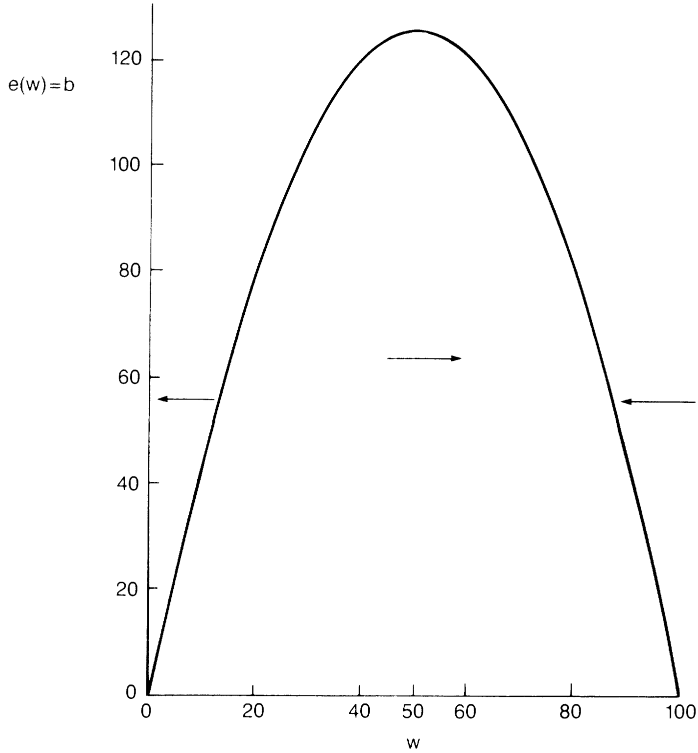


equilibrium equation (2b) for blacks. The situation is symmetrical for whites, and $e(w)$ also corresponds to equation (2a) above.[7]

Figure 2 gives the cdf for a straight-line tolerance distribution. The general equation is $F(s) = (1/R)[1 + R - (1/s)]$, where s (thresholds) is greater than $[1/(1 + R)]$, and $F(s) = 0$ otherwise; here, we let $R = 5$. Though system equations are piecewise (falling into four

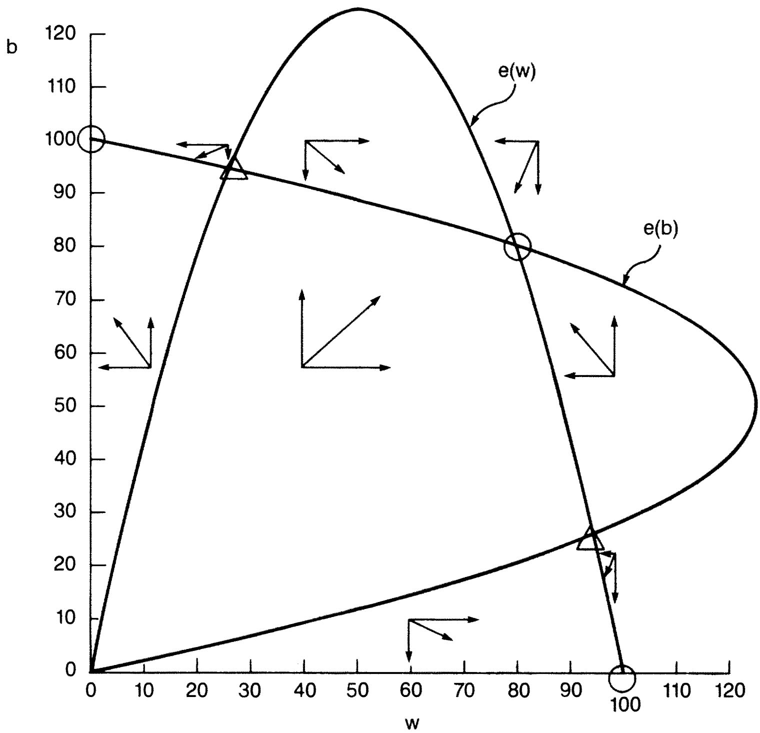
[7] Our definition of these curves is broader than Schelling's. He finds the maximum number of whites that a given number of blacks could live with. But we define $e(b)$ to be all those numbers of whites for which some given number of blacks would remain unchanged in the next time period. The two definitions are identical except when the cdf has a horizontal segment, but in the discrete distributions we would find empirically, this is common. Then the procedure of finding b/N_b on the vertical axis of the cdf (drawing a horizontal line to the cdf and dropping a line to the x axis to find the argument of F that corresponds to that proportion) no longer yields a unique solution. In such cases, $e(b)$ is not a single-valued function, because there is a range of numbers of whites for whom there would still be just b blacks in the next time period. When we later introduce preferences for integration, we will find that the definition given by the task of finding the maximum tolerable number of the other group is no longer apt in developing a graphical method but that our more general definition continues to apply.

FIGURE 3. $e(w) = b = R[1 - (w/N_w)]w$, where $R = 5$, $N_w = 100$: The equilibrium curve for whites in a straight-line tolerance distribution. Arrows indicate increases or decreases in the number of whites for given neighborhood racial compositions.



regions, as described above), $e(w)$ and $e(b)$ are continuous, since every value of w/N_w or b/N_b has a unique intersection with F. Because we can find the curve $e(w)$ explicitly by solving $w/N_w = F_w[w/(w + b)]$ for b, we substitute in the assumed equation for F to find that $e(w) = b = R[1 - (w/N_w)]w$, a parabola depicted in Figure 3 for $R = 5$, which is identical to that displayed by Schelling (1971, p. 171, fig. 19). For values of w and b within $e(w)$, more whites will enter, since the number of blacks in the neighborhood is less than the maximum that that number of whites can accept. Outside the curve, whites will leave; on the curve, no change will occur—hence, the arrows in Figure 3. The comparable parabola for blacks, with similar arrows drawn, is superimposed on the white curve in Figure 4. Within each region we indicate the resultant of the two arrows (as does Schelling 1971, p. 171). The

FIGURE 4. $e(w)$ and $e(b)$ for $R = 5$, $N_w = N_b = 100$: Equilibrium curves for whites and blacks with identical straight-line tolerance distributions. Horizontal arrows indicate direction of change in numbers of whites, vertical arrows the direction for blacks, and diagonal (resultant) arrows the direction of overall change. Stable equilibria are circled, unstable ones are enclosed in triangles.



arrows indicate that (80, 80) is a stable equilibrium, as are (0, 100) and (100, 0), and that the equilibria (25.36, 94.64) and (94.64, 25.36) are unstable, because any nearby trajectory will move away from them.

To summarize, we have graphed together the two equilibrium equations; the intersections are thus equilibria and the arrows in the phase diagram give a quick reading on stability. This method shows the dynamics clearly, which may be especially useful if the underlying equations are analytically intractable or result from empirical estimation of thresholds, as we will illustrate later. But compared to a method that works directly with the system of equations on which the graphs are based, it has two important limitations: (1) it cannot be used to indicate the details of the path to equilibrium, and (2) if there is ambiguity about the nature of stability at a particular equilibrium point, it cannot resolve it. In Figure 4, all equilibria are unequivocally stable or unstable; linearization gives eigenvalues of zero or greater than unity. But if the arrows cycled round an equilibrium point, the diagram could not tell us whether the dynamics consisted of an inward or outward spiral or some oscillation. In such a case, eigenvalue analysis would also be ambiguous, showing a value of unity. That is, this method displays graphically the same information available algebraically from linearization. When the linearization is insufficiently informative, so is the graphical method, and only by detailed analysis of the system of equations can the dynamics be clarified.

6. DECISION REVERSALS

In the model thus far, individuals make some decision only when their own group is some minimum proportion of those who have already done so. But later, if their group exceeds some still *higher* proportion, they may change their mind. Individuals who would not speak out until some minimum proportion of those expressing opinions were in their camp might no longer feel the need to speak once a more substantial proportion agreed with them and the situation seemed more securely in hand. This seems even more likely when the action in question is more costly than just expressing an opinion. In typical public-goods situations, for example, actions may require time, effort, or resources, and an individual may have a more substantial impact on the outcome as one of a few actors than as part of a large majority. For example, you may vote for your candidate only if her election is uncertain, not if she has no chance to win or if she is a shoo-in. As perceptions of your candidate's support change, your intention to vote may fluctuate accordingly.

For residential segregation, decision reversals can be interpreted as a preference for integration. Schelling's term *tolerance ratio* nicely captures the sense that there is no positive aspect to participation of the other group; they are merely tolerated up to some maximum, after which one declines to live in the neighborhood. But sometimes variety is the spice of life, and there is, up to a point, a positive preference for the presence of the other group.

We propose a simple formal argument for decision reversals: Each individual is characterized not only by a threshold for the minimum proportion making some decision who are members of his own group but also by a threshold for the maximum proportion. An individual will join in only if the proportion of his own group among those who have already done so falls between these lower and upper thresholds.[8]

A difference-equation model then follows from an argument similar to that used when only lower (segregationist) thresholds are present. Suppose that 75 percent of a neighborhood is white at time t. What proportion of the neighborhood will be white at time $t+1$? A white who would reside in such a neighborhood would have a lower threshold less than or equal to 0.75. The proportion so willing is $F_{Lw}(0.75)$, where F_{Lw} is the cdf of lower thresholds. But some of these whites may have upper thresholds below 0.75; i.e., they may be unwilling to live in a neighborhood that is as much as 75 percent white. The proportion of such whites is $F_{Uw}(0.75)$, where F_{Uw} is the cdf of upper thresholds for whites. Thus, the proportion willing to live there at $t+1$ are those for whom 75 percent white is neither too few nor too many—namely, the difference $F_{Lw}(0.75) - F_{Uw}(0.75)$, which we shall call $G_w(0.75)$. Note that G, the difference of two cdf's, is itself no longer a cdf. The system can then be described by the coupled difference equations[9]

$$n_w(t+1) = N_w G_w(p_w(t)), \tag{5a}$$

$$n_b(t+1) = N_b G_b(p_b(t)), \tag{5b}$$

where p_w and p_b again refer to the proportions of whites and blacks in the neighborhood.

Since the form of this system is exactly that of equations (1a) and (1b), stability analysis is identical, except that the f_w and f_b of equation (3) are replaced by g_w and g_b, where g is the first derivative of

[8] Schelling asserts that his bounded neighborhood model can be interpreted, without change, to reflect integrationist preferences. But he also notes that it does not accommodate a "lower limit to the acceptable proportion of opposite color, i.e., an upper limit to the proportion of like color in the neighborhood" (1971, p. 180, or 1978, p. 165). As Taylor points out, however, "[to] accurately model a preference for integration... there must be some account taken of the lower limit. A preference for integration means precisely that people will try to live in neighborhoods that are neither too white nor too black" (1984, p. 151n).

[9] A similar argument is presented for threshold models of collective behavior in Granovetter and Soong (1983) and is applied to the analysis of consumer demand in Granovetter and Soong (1986).

the difference function G. Thus, we have asymptotic stability iff

$$N_w g_w[w/(w+b)]\left[b/(w+b)^2\right] + N_b g_b[b/(w+b)]\left[w/(w+b)^2\right] \tag{6}$$

has absolute value less than unity.

We offer one example of dynamics with preferences for integration. With both segregation (lower) and integration (upper) thresholds to consider, we have a bivariate density. By definition, one's lower threshold must be lower than or equal to one's upper threshold. Thus, the density is distributed not over the entire unit square but over the upper left triangle, where lower thresholds are on the x axis and upper thresholds are on the y axis. Figure 5 displays one simple bivariate density: The distribution is uniform over a rectangular section of the triangle and zero elsewhere. Then, everyone's lower threshold is less than or equal to some parameter a, and everyone's upper threshold is greater than or equal to a. This is plausible, though it is not required by the assumption of integration preferences. It has the advantage of allowing representation of the density by one parameter. We simplify further by assigning this distribution to both groups, but with possibly different a's, indexed as a_w and a_b.

The bivariate density for whites corresponding to Figure 5 must be just that height above the rectangle that, multiplied by its area, yields the unit volume required for a density. Thus,

$$\begin{aligned} f_w(s_L, s_U) &= 1/\left[a_w(1-a_w)\right] && \text{within the rectangle and} \\ &= 0 && \text{elsewhere.} \end{aligned} \tag{7}$$

Then, the marginal densities for lower thresholds are

$$\begin{aligned} f_{Lw}(s_L) &= 1/a_w, && \text{where } 0 \le s_L \le a_w, \text{ and} \\ &= 0 && \text{elsewhere,} \end{aligned} \tag{8}$$

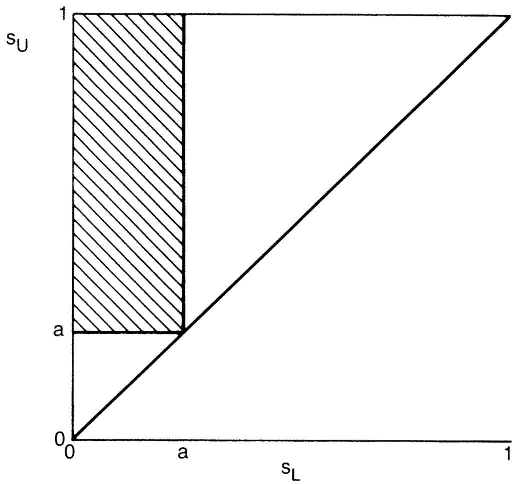
and the marginal densities for upper thresholds are

$$\begin{aligned} f_{Uw}(s_U) &= 1/(1-a_w), && \text{where } a_w \le s_U \le 1, \text{ and} \\ &= 0 && \text{elsewhere.} \end{aligned} \tag{9}$$

Corresponding marginal cdf's for lower thresholds are

$$\begin{aligned} F_{Lw}(s_L) &= s_L/a_w, && \text{where } 0 \le s_L \le a_w, \text{ and} \\ &= 1 && \text{elsewhere,} \end{aligned} \tag{10}$$

FIGURE 5. Bivariate density of upper (s_U) and lower (s_L) thresholds, uniform over shaded rectangle and zero elsewhere.



and for upper thresholds are

$$F_{Uw}(s_U) = (s_U - a_w)/(1 - a_w), \qquad \text{where } a_w \le s_U \le 1, \text{ and} \qquad (11)$$
$$= 0 \qquad \text{elsewhere.}$$

The difference, G_w, is then $F_{Lw} - F_{Uw}$, and it is given by

$$G_w(s) = s/a_w, \qquad \text{where } 0 \le s \le a_w, \text{ and} \qquad (12)$$
$$= (1 - s)/(1 - a_w), \qquad \text{where } a_w < s \le 1.$$

All these equations, with different subscripts, apply to blacks.

From (5a) and (5b) we see that for the preferences in Figure 5, system equations for whites are

$$n_w(t+1) = (N_w/a_w)p_w(t), \qquad \text{where } 0 \le p_w(t) \le a_w, \text{ and} \qquad (13)$$
$$= [N_w/(1 - a_w)]\, p_b(t), \qquad \text{where } a_w \le p_w(t) \le 1,$$

and for blacks are

$$n_b(t+1) = (N_b/a_b)p_b(t), \qquad \text{where } 0 \le p_b(t) \le a_b, \text{ and}$$
$$= [N_b/(1 - a_b)]\, p_w(t), \qquad \text{where } a_b \le p_b(t) \le 1.$$

That is, system equations are defined in piecewise fashion. For each group, the first equation applies if the group's current proportion of the neighborhood is less than or equal to its parameter a. Otherwise, the second equation applies.

Because two equations are possible for each group, there are four sets of possibilities, or parameter regions, depending on current neighborhood composition.

Conditions for region 1 are $p_w(t) \leq a_w$ and $p_b(t) \leq a_b$. Both groups are at or below their parameter a, so behavior in neither group is affected by integrationist preferences.[10] System equations are

$$\begin{aligned} n_w(t+1) &= (N_w/a_w)p_w(t), \\ n_b(t+1) &= (N_b/a_b)p_b(t). \end{aligned} \tag{14}$$

In equilibrium these reduce to

$$\begin{aligned} w &= (N_w/a_w)[w/(w+b)], \\ b &= (N_b/a_b)[b/(w+b)], \end{aligned} \tag{15}$$

whence we have $w + b = (N_w/a_w) = (N_b/a_b)$. Equilibria thus exist in this region iff this unlikely parameter condition is met. Then, any point on the line defined by $w + b$ that lies in region 1 is an equilibrium. The stability of these equilibria is assessed by using equations (6), (12), and (15) to compute the relevant eigenvalue, which then turns out to be exactly unity, which is uninformative on stability.

But inspection of the equations (14) shows that when equilibria exist, dynamics consist merely of preserving the initial proportions of whites to blacks but scaling the initial numbers up or down so they add to the constant amount $(N_w/a_w) = (N_b/a_b)$. Thus, if $N_w = 200$, $a_w = 0.8$, $N_b = 100$, and $a_b = 0.4$, then this constant amount is 250, and an initial condition of 100 whites and 50 blacks is scaled up to an equilibrium of 166.67 whites and 83.33 blacks, still in the proportion of 2 to 1. A perturbation of this equilibrium will return the system to the same equilibrium only if it preserves this proportion; otherwise, the system will move to a different equilibrium on the line $w + b = 250$, determined by the new proportions.

Conditions for region 2 are $a_w < p_w(t) \leq 1$ and $0 \leq p_b(t) \leq a_b$. In this region, some whites but no blacks are affected by preferences for integration. We then have

$$\begin{aligned} n_w(t+1) &= [N_w/(1-a_w)]\, p_b(t), \\ n_b(t+1) &= (N_b/a_b)p_b(t). \end{aligned} \tag{16}$$

[10] In region 1, we must have $(a_w + a_b) \geq 1$, since if the sum were less than 1 and if each current proportion were less than its respective a, the sum of current proportions would not equal 1, as it must.

In equilibrium, $w = [N_w/(1 - a_w)][b/(w + b)]$ and $b = (N_b/a_b)[b/(w + b)]$. Solving for w and b in terms only of the system parameters, we get

$$\begin{aligned} w &= [N_w/(1 - a_w)]/\{1 + [N_w a_b/((1 - a_w)N_b)]\}, \\ b &= (N_b/a_b)/\{1 + [N_w a_b/((1 - a_w)N_b)]\}. \end{aligned} \tag{17}$$

We may gauge stability of equilibrium by substitution into (6), using (12) and (17), yielding an eigenvalue of zero; thus, the equilibrium of equations (17) is asymptotically stable.

Conditions for region 3 are $0 \le p_w(t) \le a_w$ and $a_b < p_b(t) \le 1$. This region is exactly symmetric to region 2, interchanging the two groups. System equations and equilibria are thus identical, changing b for w, and the variable eigenvalue is again zero, indicating stable equilibria.

Conditions for region 4 are $a_w < p_w(t) \le 1$ and $a_b < p_b(t) \le 1$.[11] Here, some whites and some blacks will feel that there are too many of their own color in the neighborhood. System equations are

$$\begin{aligned} n_w(t+1) &= [N_w/(1 - a_w)]\, p_b(t), \\ n_b(t+1) &= [N_b/(1 - a_b)]\, p_w(t). \end{aligned} \tag{18}$$

Some algebra shows that the equilibria are

$$\begin{aligned} w &= [N_w/(1 - a_w)]^{1/2}/\{[(1 - a_w)N_w]^{1/2} + [(1 - a_b)/N_b]^{1/2}\}, \\ b &= [N_b/(1 - a_b)]^{1/2}/\{[(1 - a_w)/N_w]^{1/2} + [(1 - a_b)/N_b]^{1/2}\}. \end{aligned} \tag{19}$$

From (6), (12), and (19) we find that the second eigenvalue here is -1, indeterminate for stability. But if we trace out equations (18) on a calculator, we see that for any set of parameters, every set of initial conditions $n_w(t)$ and $n_b(t)$ leads immediately to an oscillation of period two. The equilibrium point given by the parameters is within the range of the oscillation but is itself never reached. Further, each set of initial conditions leads to its own particular oscillation, unique except that every pair in the same ratio leads to the same oscillation. Finally, the initial white/black ratio reappears every second period. These characteristics could be derived analytically by computing general expres-

[11] In region 4, we must have $(a_w + a_b) \le 1$, since both groups exceed their a parameter. This could not be the case if the sum of the two parameters exceeded 1.

sions for $n_w(t+2)$ and $n_b(t+2)$, but it is more illuminating to give a numerical example.

Suppose that a neighborhood has 200 whites and 100 blacks and that $a_w = 0.3$, $a_b = 0.1$, $N_w = 300$, and $N_b = 300$. Each group contains some individuals who believe that there are too many of their own group present. From (19) we see that an equilibrium exists at $w = 200.84$, $b = 177.12$. But from (18) we can compute that at time $(t+1)$, the numbers of whites and blacks are $(142.86, 222.22)$; at $t+2$ we have $(260.87, 130.43)$, which reproduces the original 2-to-1 ratio; at $t+3$ we return to $(142.86, 222.22)$, and so on. The equilibria in region 4 do not attract nearby trajectories. If we were at the equilibrium $(200.84, 177.12)$ and were slightly pushed away to, for example, $(201, 177)$, the system would not return to the equilibrium but would not move very far. Rather, it would settle into the cycle $(200.68, 177.25)$, $(201, 177)$. Similarly, any cycle in progress, if slightly perturbed, would settle into a close-by cycle. Thus, the initial conditions under which the system first enters region 4 determine the amplitude of oscillation.

While the system is entirely deterministic, it is, in two of the four regions, enormously sensitive to initial conditions and may fluctuate in apparently odd ways even without perturbations. Such fluctuations are characteristic of systems of deterministic nonlinear difference equations, and May (1976) has shown that under certain conditions, they may be essentially indistinguishable from random noise. In the general case, numerical methods are required to determine outcomes, but some analytical approximations to observed distributions are usually possible, and these can be related to the techniques used here, to give a reasonable picture of system dynamics.

7. THE GRAPHICAL METHOD WITH DECISION REVERSALS

The graphical method described earlier can also be applied when decisions are reversible. In the segregation example, we again ask, Given a certain number of blacks in a neighborhood, for what number of whites in the neighborhood will that number of blacks not change? When integration preferences are present, there may be more than one such number of whites, even if there are no horizontal segments in the cdf, because there may now be not only too many whites for the fixed number of blacks but also too few.

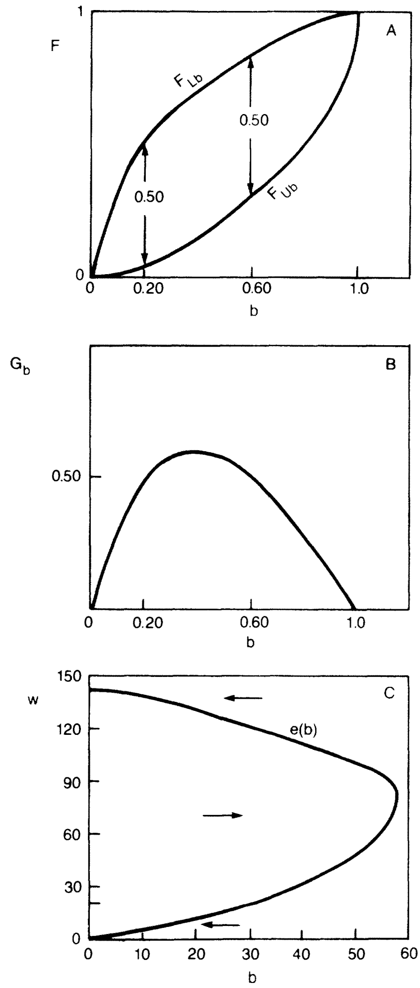
Suppose we ask, For what numbers of whites will a neighborhood that currently has 30 of a population of 60 blacks continue to have just 30 blacks? The basic difference equation (5b) may be written $n_b(t+1)/N_b = G_b(p_b(t))$, where p_b is the proportion of the neighborhood that is black; i.e., $b(t)/(w(t)+b(t))$. Since the 30 blacks are just 50 percent of the black population, if the neighborhood is to continue to have 30 blacks at $t+1$, the 30 must be a proportion of the neighborhood equal to p_b such that p_b satisfies the relation $0.50 = F_{Lb}(p_b) - F_{Ub}(p_b) = G_b(p_b)$. This is the equilibrium condition for the equation. But since G need not be monotonic, more than one value of p_b may satisfy this relation; i.e., there may be more than one proportion of whites in the neighborhood such that 50 percent of the black population will continue to live in the neighborhood. Any p_b for which the difference in height between lower and upper threshold cdf's is exactly 50 percent will suffice.

Suppose our cdf's were such that $G_b(0.60) = G_b(0.20) = 0.50$. Then, 30 blacks could coexist with exactly 20 whites and with exactly 120 whites. (Note that the 30 blacks in the first and second cases might not be the same.) Since the cdf's whose difference is being taken are monotonically increasing, numbers of whites below 20 would be too few, and some blacks would depart. At 20 whites, 30 blacks would be present next period. Between 20 and 120 whites would be neither too few nor too many, so the 30 blacks would remain and more would enter. At 120 whites, exactly 30 blacks would remain in any subsequent period, but more than 120 whites would be too many and the number of blacks would again fall below 30. If more than two arguments of G yielded the result of 50 percent, these alternations would occur once again.

Figures 6a and 6b show the cdf's and the difference function G discussed above. Our procedure for constructing the curve $e(b)$ is analogous to that outlined earlier, when preferences were only for segregation. We convert the number of individuals in the group in question—here, 30 blacks—to a proportion of the group's population—here, 50 percent—and locate this proportion on the vertical axis of the G curve. Then, we move across the curve horizontally, finding all intersections with G, and take as $e(b)$ the arguments of G corresponding to those intersections, as shown in Figure 3c. Here $e(b)$ is a vector-valued function of b.

In equations, $e(b)$ is found by solving $b/N_b = G_b(b/(w+b))$, the equilibrium equation for blacks, for w. We find $e(w)$ in the same

FIGURE 6. *A* and *B*, upper and lower threshold cdf's for blacks (F_{Ub} and F_{Lb}) such that $G_b(0.20) = G_b(0.60) = 0.50$. *C*, $e(b)$, the equilibrium curve for blacks implied by the curves in *A* and *B*. Arrows indicate direction of change in numbers of blacks.



way. The arrows in Figure 6c indicate the changes to be expected in numbers of blacks at each point in the next period. If we superimpose the curves $e(w)$ and $e(b)$, system equilibria will be at the intersections; we can then draw resultants of black and white arrows to produce a phase diagram that gives some insight into the stability of equilibria. The observations made earlier on the values and limitations of such diagrams apply here; they would not, for example, elucidate the dynamics in regions 1 and 4 of the rectangular density discussed above.

8. GENERALIZATION BEYOND TWO GROUPS

Binary decisions may involve more than two groups. For example, there may be more than two candidates to support in an election. Likewise, residential segregation may involve three or more ethnic groups. We analyze exactly three groups and obtain the essentials for generalization to any larger number. Continuing the segregation notation, we retain the subscripts w and b for whites and blacks and add a subscript h to represent Hispanics. We include integration preferences, since their absence is then merely a special case in which no individuals have upper thresholds, i.e., $F_{Uw} = 0$ so that $G_w = F_{Lw}$, etc. The natural generalization of equations (5a) and (5b) is

$$n_w(t+1) = N_w G_w(p_w(t)), \tag{20a}$$

$$n_b(t+1) = N_b G_b(p_b(t)), \tag{20b}$$

$$n_h(t+1) = N_h G_h(p_h(t)), \tag{20c}$$

where $p_i(t) = n_i(t)/[n_w(t) + n_b(t) + n_h(t)]$, the proportion of the neighborhood made up of group i at time t. Generalization to k groups is then straightforward. Analysis of the matrix of first-order partial derivatives shows that at least one eigenvalue is zero. In general, for k groups, stability will depend on $k-1$ eigenvalues, and stability requires that the absolute value of all $k-1$ must be less than unity. Stability analysis is thus much more complex for three or more groups than for two.

Equations (20) assume that each group is sensitive only to the proportion of its *own* members in the population, that it makes no distinctions among other groups and simply lumps them together in the denominator. More complex intergroup preferences could be incorporated by assuming a matrix of weights to be inserted in the denominator of the right side of the equations. Each group has one row in the matrix indicating how it weights each other group. The main diagonal can be standardized to unity. Thus, for example, the row for blacks (bw, bb, bh) might be $(3, 1, 1/2)$. This implies that for blacks, whites in the neighborhood loom larger than life, that seeing one white is three times more salient than seeing one black. For segregation thresholds (lower thresholds), this means that the tolerance of blacks for whites is lower than it would be if the weights were equal. The weight of 1/2 for Hispanics, less than unity, indicates that blacks believe Hispanics are even easier to live with than other blacks and thus, in this restricted sense, are six times more desirable than whites. For segregation thresholds, this means that when Hispanics live in the neighborhood, blacks are comfortable with fewer other blacks than they would be if all weights were equal. For upper or integration thresholds, the effect is opposite: Because whites are more salient than blacks, it takes more blacks in the neighborhood to trigger blacks' upper threshold than it would if all weights were equal; therefore, blacks stay longer. Correspondingly, since Hispanics are less salient than blacks, their presence causes blacks to reach their upper threshold sooner.[12]

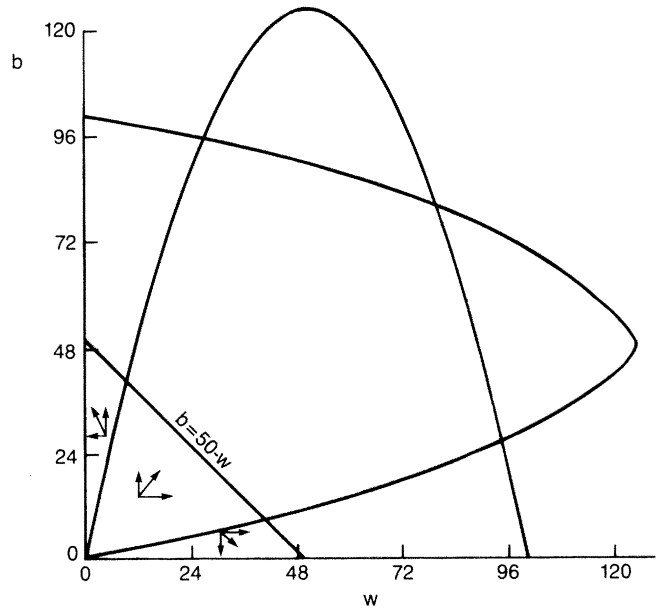
9. CAPACITY CONSTRAINTS

The problem with the segregation interpretation of our models is that there is, at least in the short run, a fixed number of housing units, which may be less than the number of people who want to live in the neighborhood. In neoclassical economic equilibrium, housing prices would adjust to equate supply with demand; but housing markets are often out of equilibrium, so it is useful to indicate how our model is affected.

Consider the neighborhood in Figure 7: Demand comes from 100 whites and 100 blacks, and stable equilibria are (80,80), (100,0), and (0,100). Suppose there are just 50 homes. This capacity constraint can be indicated by drawing in the line $b = 50 - w$, as indicated (see Taylor [1984, p. 146] for a similar diagram). Suppose thresholds and dynamics are as given by equations (4) and that we begin with 10

[12] Weights could also be assigned in the analysis of two groups, but these would then simply alter the unweighted thresholds. For example, if whites weighted blacks three times more than whites, a white who had an unweighted threshold of 0.5 would live in the neighborhood if $n_w/(3n_b + n_w) > 0.5$, i.e., if $n_w/n_b > 3$, which is the same as a threshold of 0.75. More generally, if weight w is assigned to the other group and unity is assigned to one's own, an individual with a stated threshold of t will have a normalized or adjusted threshold of $wt/[t(w-1)+1]$. Thus, for two groups, the weights may simply be determinants of thresholds that have already been normalized for weighting considerations. However, this method fails when there are more than two groups, because no unique normalized threshold can be obtained from an initial stated threshold and a set of weights.

FIGURE 7. The curves of Figure 4, with added capacity constraints of maximum neighborhood size = 50, indicated by the line $b = 50 - w$.



whites and 20 blacks in the neighborhood. This places us in region 4, so equations (4g) and (4h) indicate what will occur. Recalling that $R = 5$, we see from these equations that at time $t + 1$, 60 whites and 90 blacks will want to live in the neighborhood. Since this exceeds the stated capacity of 50, some rationing rule is required. One, but not the only, plausible rule is to fill the houses with whites and blacks according to their proportions among the candidates. Then, we would have 20 whites and 30 blacks in the next period. What will happen next? Applying equations (4g) and (4h) again shows that at $t + 1$, 70 whites and 86.7 blacks will want to live in the neighborhood. If we again take 50 whites and blacks according to their proportions among those who want to live in the neighborhood, we have 22.33 whites and 27.67 blacks. Continuing this process leads us closer and closer to a limiting value of 25 whites and 25 blacks, because the equilibrium, given a capacity constraint and the rationing rule adopted here, is simply the unconstrained equilibrium $(80, 80)$ scaled down linearly to meet the constraint. The stability characteristics of unconstrained equilibria are identical to those of the constrained equilibria.[13]

But there is a troublesome aspect to these dynamics. With 20 whites and 30 blacks in the neighborhood, we applied the system equations and the rationing rule to find that 70 whites and 86.7 blacks would be candidates for the neighborhood in the next period, and thus 22.33 whites and 27.66 blacks would live there. But the 70 and 86.7 include the 20 whites and 30 blacks already in the neighborhood; therefore, our procedure requires some resident blacks to leave even though they are *not* dissatisfied. This complete reshuffling each time period makes sense only when transaction costs of entering or leaving a state are negligible, which is implausible for real neighborhoods. An additional stipulation, which modifies the dynamics of equations (4), seems necessary: Residents in a neighborhood will not vacate if existing racial proportions are such that they would like to live there next period. In the model without capacity constraints, this question cannot arise, since anyone who wants to live in the neighborhood can.

Then, 20 whites and 30 blacks represents an equilibrium, since no one is motivated to leave. In fact, every point on the line $b = 50 - w$ that is within the two parabolas is an equilibrium. These equilibria have only limited stability, however. Any perturbation that leaves whites and blacks in the ratio of 2/3 will restore the 20/30 outcome. But any other change will result in a new equilibrium, along the line. In particular, if a perturbation results in a ratio of x whites to y blacks, the new equilibrium point will be that which scales x and y up to the line, i.e., to that point (w, b) at which $w/b = x/y$ and $w + b = 50$. As the arrows in Figure 7 suggest, points on the line $b = 50 - w$ that are not within both curves lead to either $w = 50$ or $b = 50$, which are stable against perturbations that remain outside one of the curves.

Thus, even with capacity constraints, the basic model works when demand for places does not exceed supply—market equilibrium being included as the special case of equality—or when transaction costs of entering or leaving are negligible. When neither condition is met, system outcomes will be affected in important ways that must be taken into account. In the public opinion interpretation, capacity constraints are likely to be irrelevant, since there is no obvious upper

[13] These assertions seem intuitively reasonable to us, so we omit the proofs. However, they are available on request.

limit to the number of individuals who can express their views. Club membership usually involves a capacity constraint, but the transaction costs of entering and leaving may be minimal. Sensing the volatility this implies, many clubs thus impose nonrefundable initiation fees.

10. EMPIRICAL APPLICATIONS

Empirical applications require measurement of the postulated thresholds. We see two ways to proceed. One way is to adapt the economists' concept of revealed preference, attributing thresholds to individuals by observing the distribution of others' decisions before they make their own. This has the advantage of resting on observed facts, but for any given individual, observation is uninformative unless he actually does something. A black who remains in his neighborhood as it changes from 100 percent black to 60 percent is known to have a threshold of 60 percent or less, but this right-censored observation is crude. When we take upper thresholds (integration preferences) into account, it becomes clear that good measurement requires observation over a rather wide range of neighborhood conditions. We must also assume that behavior exactly reflects thresholds and that there is no significant lag between the passing of one's threshold and behavior. If there are lags, and worse, if these vary by individual, then the composition of the neighborhood or of expressed opinion just before movement or the expression of one's views may not accurately reflect the actor's threshold.

The other way to measure thresholds is to ask respondents directly. This method is suspect to the extent that there is no independent check on the validity of respondent reports. However, because it is direct, it does not suffer from the censoring and lag problems of revealed preference measures, and it appears empirically that respondents have no difficulty answering questions of this kind. Noelle-Neumann has shown that peoples' willingness to express their opinions depends on whether they think they are in the majority. She has also explored other determinants of this willingness (1984, chap. 2). She asks respondents, for example, whether, faced with a five-hour train ride with a stranger who has one of two definite views on some subject, they think it would be worth their while to discuss the subject with that person (1984, p. 18). We could extend this question and ask respondents how interested they would be in expressing their opinion in some public setting in which the group was split in various proportions on some question.

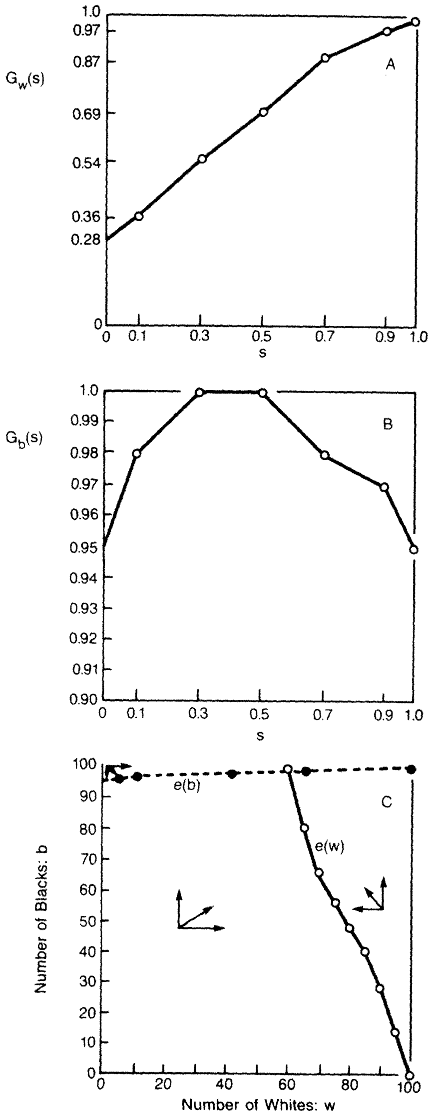
For the interpretation of residential segregation, Taylor (1984) reports successful use of a question like this in a telephone survey of 300 black and 300 white residents of Omaha in 1978 on the importance of racial preference in housing choice. Neighborhoods were described as having certain numbers of blacks and whites, and respondents were asked whether they would try to move out of their neighborhood if its racial proportions changed to those given in the question (Taylor 1984, p. 244, n. 1). Each respondent was asked about seven possible mixtures: all white, 1 black/9 white, 3 black/7 white, 5 black/5 white, 7 black/3 white, 9 black/1 white, and all black. Though Taylor drops that part of the data reflecting preferences for integration (1984, p. 151), the question, since it captures for each respondent the acceptable range of neighborhood compositions, does index both segregation and integration preferences and thus permits us to estimate both lower and upper threshold distributions. For example, a white respondent who accepts black/white ratios of 9/1, 7/3, 5/5, and 3/7 but who rejects ratios of 10/0 and 1/9 has a lower threshold of 10 percent and an upper threshold of 90 percent. He is willing to live in a neighborhood that is between 10 percent and 90 percent white. In Table 1 we give Taylor's empirical results, and in Figures 8a and 8b we estimate the corresponding functions $G_w(s)$ and $G_b(s)$ by connecting the measured points with straight lines. (Other assumptions

TABLE 1
Tolerance Schedules Estimated from Omaha Survey Data

Racial Mix	Percentage Who Would Tolerate Mix: White	Percentage Who Would Tolerate Mix: Black
All white	100	95
1 black, 9 white	97	98
3 black, 7 white	87	100
5 black, 5 white	69	100
7 black, 3 white	54	98
9 black, 1 white	36	97
All black	28	95

Source: Taylor 1984, p. 150.

FIGURE 8. *A*, $G_w(s)$: The white difference curve interpolated from the data points of Table 1. *B*, $G_b(s)$: The black difference curve interpolated from the data points of Table 1. *C*, $e(w)$ and $e(b)$ for $N_w = N_b = 100$: Equilibrium curves for whites and blacks implied by the curves in *A* and *B*.



for interpolating intermediate points are possible, but this seemed the most straightforward.) The function $G_w(s)$ is identical to a cdf, $F_w(s)$, of lower (segregation) thresholds, since no whites in this survey indicated any preference for integration.

To construct the equilibrium curves $e(w)$ and $e(b)$, we apply the graphical method described above. For a given number of blacks, for example, we set $b = N_b G(p_b)$ and solve to find w, the point(s) on the curve $e(b)$. We need only specify the numbers of whites and blacks (N_w and N_b) interested in living in a given neighborhood. Taylor draws these curves for 100 of each race (1984, p. 153) but excludes preferences for integration, since these were not treated by Schelling. Our Figure 8c gives $e(b)$ and $e(w)$, including integration preferences. A stable equilibrium point occurs at about 99 blacks and 60 whites, which is 62.3 percent black and 37.7 percent white. Without the integration preferences, $e(b)$ would be a horizontal line at $b = 100$, indicating that for any number of whites from 0 to 100, all 100 blacks in the neighborhood at time t would remain at $t + 1$. The curves would then cross at 100 blacks and about 59 whites. In this case, integration preferences do not make much difference, since no whites have them and only about 5 percent of blacks do. Notice the direction of the small difference: One black leaves, finding the neighborhood "too black," and he is replaced by one white.

Different numerical assumptions would yield very different results. Noting that Omaha was about 12 percent black at the time of the survey, suppose we had 880 whites and 120 blacks interested in some neighborhood. The intersection of $e(w)$ and $e(b)$ occurs at about 118 blacks and 844 whites; whites are then 87.7 percent of the neighborhood, almost exactly the same as their proportion in the population.

11. CONCLUSIONS

Our models fit cases in which one's behavior depends on the previous behavior of others and in which the group composition of those making a certain decision is the crucial element of that dependence. One application is residential segregation. We showed that Thomas Schelling's models can be expressed in terms of our system of two coupled difference equations, permitting an exact mathematical account of his results and such important generalizations as the incorporation of preferences for integration and the extension to more than two groups. The model applies also to a class of public-opinion problems, usually discussed under the rubric of pluralistic ignorance. Another natural linkage is to free-rider problems and the provision of public goods. The techniques of Noelle-Neumann (1984) and Taylor (1984, 1986) suggest that thresholds can be measured with relatively straightforward techniques of survey analysis.

These models have three distinct advantages over most current models: (1) their treatment of dynamics is explicit and central (i.e., they do not deal in comparative statics), (2) they make no assumptions of linear relations among variables, and (3) they are driven not by correlations but by well-defined causal mechanisms. We see models of this kind as part of a broader movement in sociology toward explicit, concrete, dynamic analysis and away from the general linear model, which, assuming that the size of causes must determine the size of consequences, prepares us poorly for the many surprises that social life has in store.

REFERENCES

Granovetter, Mark. 1978. "Threshold Models of Collective Behavior." *American Journal of Sociology* 83:1420–43.

Granovetter, Mark, and Roland Soong. 1983. "Threshold Models of Diffusion and Collective Behavior." *Journal of Mathematical Sociology* 9:165–79.

______________. 1986. "Threshold Models of Interpersonal Effects in Consumer Demand." *Journal of Economic Behavior and Organization* 7:83–99.

Luenberger, David. 1979. *Introduction to Dynamic Systems*. New York: Wiley.

May, Robert. 1976. "Simple Mathematical Models with Very Complicated Dynamics." *Nature* 261:459–67.

Noelle-Nuemann, Elisabeth. 1974. "The Spiral of Silence: A Theory of Public Opinion." *Journal of Communication* 24:43–51.

______________. 1977. "Turbulences in the Climate of Opinion: Methodological Applications of the Spiral of Silence Theory." *Public Opinion Quarterly* 41:143–58.

______________. 1984. *The Spiral of Silence*. Chicago: University of Chicago Press.

O'Gorman, Hubert, and Stephen Garry. 1976. "Pluralistic Ignorance: A Replication and Extension." *Public Opinion Quarterly* 40:449–58.

Schelling, Thomas. 1971. "Dynamic Models of Segregation." *Journal of Mathematical Sociology* 1:143–86.

______________. 1973. "A Process of Residential Segregation: Neighborhood Tipping." Pp. 157–84 in *Racial Discrimination in Economic Life*, edited by Anthony Pascal. Lexington, MA: D. C. Heath.

________________. 1978. *Micromotives and Macrobehavior*. New York: W. W. Norton.

Spence, Michael. 1974. *Market Signaling*. Cambridge, MA: Harvard University Press.

Taylor, D. Garth. 1982. "Pluralistic Ignorance and the Spiral of Silence: A Formal Analysis." *Public Opinion Quarterly* 46:311–35.

________________. 1983. "Public Opinion and Community Conflict: Threshold Models, The Spiral of Silence and Anti-Busing Protest." Mimeo.

________________. 1984. "A Revised Theory of Racial Tipping." Pp. 142–66 in *Paths of Neighborhood Change: Race and Crime in Urban America*, edited by Richard Taub, D. Garth Taylor, and Jan Dunham. Chicago: University of Chicago Press.

________________. 1986. *Public Opinion, Collective Action and Anti-Busing Protest: The Boston School Desegregation Conflict*. Chicago: University of Chicago Press.