

Formal Models in Studying Collective Action and Social Movements
Pamela E. Oliver
Daniel J. Myers

Preprint of chapter in *Methods of Research in Social Movements*, Bert Klandermans and Suzanne Staggenborg, editors. University of Minnesota Press. (2002)

Introduction

Our purpose in this chapter is to explain the value of mathematical modeling of social movement phenomena. We have the daunting task of speaking to two very different audiences: those comfortable with mathematics who want to develop mathematical models relevant to social movements, and those uncomfortable with mathematics who want to learn helpful things about movements from the mathematics-based work of others. We will try to speak to both audiences, and we ask the forbearance of each toward those sections that aimed at the other.

For those who begin feeling that mathematics is an alien and even dehumanizing tool, we suggest that mathematical language can be understood in cultural terms, as a mode of communication that uses particular symbols and patterns to convey meaning. Like any unfamiliar culture, the modes of expression seem alien at first, but when you know the language and its meanings, you can recognize its beauty and discover that some ideas can be expressed more clearly in that language than any other. Mathematics is a language that permits thoughts and new ideas that simply cannot be expressed in other ways. If you don't speak the language of mathematics, it is like reading poetry in translation: you get some of the ideas, but never the full power and beauty of the original.

To expand this point further, consider an example from statistics (not formal models), but one familiar to most sociologists: a table of multivariate regression coefficients. Such a table summarizes information that would otherwise require several pages to explain and does so in a format that is easier to understand and evaluate than a verbal description ever could be. You do need to learn the cultural practice or language of a regression table to understand it, but if you know the language, it is a very efficient and clear mode of communication. Knowing the language of a regression table is essential for sociologists, and so is knowing enough mathematical notation to read basic algebraic equations, even if you do not want to do mathematical sociology yourself.

The translation metaphor also has meaning for those who do "speak" mathematics. There is value in carefully translating mathematics into words, so that those who do not readily grasp equations can appreciate the ideas they convey. And, just as some have been inspired to learn Italian or Arabic to appreciate Dante or the Koran, some sociologists have been inspired to learn more mathematics from the interest sparked by verbal translations of mathematical sociology.

In this spirit, we welcome all of you to a discussion of mathematical models in the study of collective behavior and social movements. We approach this enterprise with an assumption that the full understanding of any social phenomenon requires many different approaches and methodologies, and that our task is to explain the value of mathematical models as one of them. Analogies to poetry notwithstanding, we still have the language problem. All but the very simplest mathematical formulations cannot be explained without equations, and a full exposition

of any complex model can require twenty pages or more. Just as the chapters about survey research or participant observation cannot actually let you experience the research method, we are reduced to writing about mathematical methods while only demonstrating them on a cursory level.

The Problem of Collective Action

Let's begin where math models in social movements began, with the problem of collective action. Although "collective action" might refer to anything people do together, most social scientists have defined it as an action which provides a shared good, deriving their view from Mancur Olson's The Logic of Collective Action (1965). Olson woke up social science by claiming that "rational, self-interested individuals will not act to achieve their common or group interests" (1965:2). Prior to Olson, sociologists assumed a natural tendency led people to act on shared interests. But economists had long argued that coercive taxation is necessary because rational individuals would not voluntarily contribute to public goods such as armies, public schools, or sewage systems which could not be withheld from those who did not pay. Olson argued that all group interests were subject to the same dilemma because when benefits cannot be withheld from non-contributors, rational individuals are motivated to free ride on the contributions of others. Olson influenced early resource mobilization theory by focusing attention on the problem of getting people to participate in collective action, and thus, the last thirty years of social movements theory.

Olson accompanied his verbal arguments with equations. By the 1970s, however, many had argued that Olson's equations were too limited for use in further theorizing and began to develop other ways of expressing the problem. In the process, they stopped asking if collective action is rational and began identifying conditions where collective action was more or less likely (see Hardin 1982; and Oliver 1993 for reviews).

One line connected Olson's problem to the Prisoners' Dilemma (PD). The original "story" of the PD game is that two criminals are caught for committing a burglary together and interrogated separately. If neither confesses (both cooperate), each will get a two year sentence. If both confess (both defect) each will get a 6 year sentence. If one confesses (defects) while the other does not, the defector gets immunity while the cooperator gets ten years. The PD game reflects Olson's problem because both players are tempted to "free ride" on the cooperation of the other given that each always benefit from defecting. Hardin (1971) argued that collective action was a prisoners' dilemma between "self" and "the group," and the PD tradition continues to be a major framework for analysis. However, others (e.g. Runge 1984, Cortazar 1997, Hamburger 1979, Chong 1991, Heckathorn 1996) argued that collective action can also be an "assurance game" in which all benefit if all cooperate, but are hurt if someone defects. Game theory provides a rich history of considering the strategies derived from various payoff structures, rules about repeating the game, and how players communicate. But while this tradition is useful for analyzing strategic interaction in two-actor systems and certain small group situations, it is too cumbersome and intractable for modeling action in large heterogeneous groups.

The approach that has proved more flexible in the long run begins with decision theory equations. Decision equations are based on the idea that people will do things that bring them net benefits. Theorists have developed various ways of expressing an individual's benefits as a function of his/her own actions and the actions of others in the group. To translate this kind of

idea into mathematics, we will identify five important elements. First, there is the outcome of the model we'll call G_i , which is the net gain to any individual i . The second element is the costs of contributions to the collective good. We'll allow individual contributions of different sizes and call a contribution size r , and call C_i the cost to the individual of making a contribution of size r . In addition, we will represent the contributions of all others as R . The third element, P , is the amount of the collective good that is provided (to everyone) based on the total contribution ($R+r$ if i contributes, and R if i does not contribute). Of course, levels of provision have different value to each individual, so we will call v_i the value of P to an individual. Finally, we allow for selective incentives, I , which is the value of any private incentives given to contributors. Following Oliver (1980) then, these elements are combined into a general model:

$$G_i(r) = v_i[P(R + r) - P(R)] + I - C_i(r),$$

in which the net gain to i is a function of the value v_i accruing from their contribution level r plus selective incentives, minus the costs incurred.

We then need a rule for how behavior is affected by the net payoff $G_i(r)$. A "determinate" decision rule common in economics says that a person will choose the action with the highest payoff, regardless of whether contributing versus not changes the payoff by 1 unit or 100 units. Psychologists predict behavior more probabilistically. If the payoff difference between contributing and not is 51 vs. 49, they would predict that actors would contribute 51% of the time and not contribute 49% of the time, while if the payoff from contributing is 95 and of not contributing is 5, actors would contribute 95% of the time and withhold only 5%. By contrast, a determinate model would predict contributing 100% of the time in both cases because it is the option with the highest payoff. Determinate decision rules are easier represent and manipulate in equations, while probabilistic decision rules usually fit empirical data better.

Using the simpler determinate decision model, we can predict cooperation if the net payoff is greater than zero, i.e. $G_i(r) > 0$. We can use elementary algebra to show that $G > 0$ if $[P(R+r)-P(R)] > (C_i(r) - I) / v_i$. The term $[P(R+r)-P(R)]$ is a production function, which gives the difference in the payoff P produced by a contribution r . If r makes no difference in P (as Olson argued), this term will be zero, and no level of contribution is ever rational unless $0 > (C_i(r) - I) / v_i$ which is true only if the selective incentive is greater than the cost ($I > C_i(r)$). This is exactly the situation Olson had in mind; the collective good (P) makes no difference in the outcome, *only* the relation between the cost and the incentive. However, if $P(R+r) > P(R)$ then collective action might be rational without incentives, depending on the cost.

There are several important aspects of this example that recur in models of collective action, and mathematical models more generally. First, it lays out a clear way of talking about the problem and identifying what factors are to be considered. Second, standard mathematical rules, in this case algebra, allow us to derive new relations from the given information. Third, while the new derivations were completely present in the original equation, they may not have been obvious until we performed the manipulation. Finally, and most importantly, we cannot "solve" this equation to determine if collective action is rational. The most important factor is whether r makes a difference in P , and nothing in the equation tells us whether or not that is so. We will have to make some additional assumptions about the nature of that relationship to get an answer. Theorists have had spirited arguments about which assumptions are reasonable and the conditions under which different assumptions are reasonable, but there is no mathematical proof that can resolve the matter. These untested assumptions are called the "scope conditions" of a theory, and we will say more about them below.

Determinate individual decision models are also used as parts of more complex models of interdependent decisions involving many heterogeneous individuals. Modeling multiple actors requires developing additional rules for how their actions affect each other. Oliver et al (1985) assumed that people make decisions sequentially, and showed that heterogeneous groups would behave differently from single individuals or homogeneous groups, depending on the shape of the production function P . They emphasized the "critical mass," the subset of actors with high interest in the collective good who play special roles in collective action. In some cases, the production function is decelerating so that the difference $P(R+r)-P(R)$ gets smaller as the total number of prior contributions increases: in this case, the critical mass provides the good while everyone else free rides. In other cases, the production function is accelerating so that the difference $P(R+r)-P(R)$ gets larger as the total number of prior contributions increases: in this case, the critical mass overcomes start-up costs and creates conditions which motivate the rest to participate. They also argued (Oliver and Marwell 1988) that the relationship between a group's size and the rationality of collective action varies depending on the production function, critiquing Olson's (1965) claims that large groups could not provide collective goods.

Many scholars have taken up the problem of collective action, and we can mention only a few examples here. Macy (1990, 1991a, 1991b, 1993, 1997) has focused on changing the core decision rule from determinate to a probabilistic model of adaptive learning. Macy assumes that the baseline is not a no-cost zero point, but an aversive situation that motivates actors to try other options. Collective action happens when several actors probabilistically try cooperation together, and produce better outcomes. Heckathorn (1988, 1990, 1993, 1996) has developed models in which actors can coerce each other into cooperating. A wide variety of outcomes can occur depending on the configuration of payoffs and incentive systems. In two interesting results, he shows that "hypocritical cooperation" (making others cooperate while you privately defect) can generate collective action, and that some situations create an "altruist's dilemma" in which those who do what is good for others cause worse overall outcomes compared to people who behave selfishly. Kim and Bearman (1997) assume that interests can change, rather than remain fixed. People will change their interests, and expect others to change as well, when they encounter cooperators who have higher interests and defector with lower interests.

Although collective action theory is often called "rational action" theory, theorists have often developed models which modify the assumption of self-interested egoism. One example is Gould (1993), who assumes that individuals are motivated by fairness norms. He assumes that people neither like being exploited nor wish to be viewed as exploiters, so they adjust their contributions to match others. These fairness rules lead to cascades of adjustment until a steady-state equilibrium is reached in which everyone's fairness norm is satisfied. Although his model is too complex to completely explain here, the first step gives the core notion of the approach. In the model, one person starts contributing independent of others and everyone else begins trying to "match" their contribution to the everyone else's average contribution. As the contribution levels move away from zero, they are governed by this equation:

$$c_i(t) = \lambda \frac{\sum_{j=1}^N c_j(t-1)}{N-1}, i \neq j$$

This equation says that i 's contribution at time t equals the average of everyone else's contributions at time $t-1$ multiplied by λ , a parameter that ranges between 0 and 1, where 1 means you match the average perfectly, and 0 means you stay at zero no matter what else others do. Using this model, Gould examines the effects of network density and the position within the network of initial contributors by assuming that the fairness equation above considers only those people to whom an actor has network ties.

Notice that we cannot directly test the model to see if Gould's "fairness maximizer" assumption is better than the "self-interested egoist" assumption. Instead, different theorists make different assumptions about the core principal by which people make decisions, and then derive the consequences of those assumptions. Most sociologists would agree that different people operate under different principles, and that the same people operate under different principles in different settings. Thus theories with different core assumptions should not be evaluated as "right" or "wrong," but as more or less applicable to different situations.

GENERATING AND ANALYZING MATHEMATICAL MODELS

In the balance of this chapter, we talk about some of the fundamental principles in generating and analyzing a formal mathematical model. We outline the steps to model building and discuss some of the issues involved in each. In the process, we summarize some published works which illustrate the issues and give some brief examples from our own ongoing work in modeling the diffusion of protest and collective violence. Readers who are more interested in reading and evaluating models than in writing them should find our discussion to serve as a solid pointer to issues to consider in evaluating others' models.

1. Acquire knowledge about the process you want to model. Before developing a model of social process, it is critical to know as much as possible about the process. You should understand both the prior theory that has been done in your area, as well as be familiar relevant empirical patterns and with previous theorizing (verbal or formal) related to the process. For example, a growing body of published data shows counts of protest or collective violence events over time. We note several empirical generalizations about these data plots: 1) they tend to be wave-like, that is they go up and then come down; 2) they tend to be fairly peaked or "spiky," rising and falling much more rapidly than most continuous mathematical functions; 3) they exhibit smaller waves within larger waves. Another example is that repression sometimes suppresses protest and other times spurs more protest. Most scholars believe that the effect of repression on protest is curvilinear: moderate repression spurs protest while severe repression destroys it and zero repression makes it look unimportant. Elegant mathematical models may be published in technical journals, but they will not have impact on sociology unless they are well linked to broader theoretical and empirical issues. Knowing the literature will also help you avoid "reinventing the wheel" and produce a real advance by building on the work of others.

Finally we want to emphasize the importance of looking to other disciplines and past work for mathematical forms that may be useful. For example, Macy (1990) adapted a standard learning theory model for his analysis of collective action, while Chong (1991) adapted an economics model of supply and demand. Oliver and Myers (1998) suggest that the interdependent diffusion of collective action and regime responses might adapt biological models of the coevolution of species. Having some familiar terrain in your models will make it easier to understand their behavior and make them more stable. As much as possible, build models using

standard mathematical forms with well-known properties.

2. *Clearly specify the kind of problem you wish to solve.* Supposing that you want to explain the rise and fall of protest over time. Conceptually, there are two general approaches. The first takes one or more well-defined empirical instances for which there is data and attempts to create a mathematical model that fits the data well. Such models are widely used in engineering and physical sciences to represent specific physical systems such as manufacturing production processes or predator-prey relationships. Once the model is constructed, it can be used to determine how the outcomes of the system would change if some elements in the system were altered. Demographers use this approach to construct models of populations, showing how they change in response to a particular influence, such as an increase in contraceptive use or an increase in AIDS infection. Once a model of one particular instance has been created, it can be modified to represent other similar processes. Empirical models, therefore, can be aggregated to develop more abstract theory. In our work modeling protest cycles, we have found that the best type of mathematic model for representing the basic "look" of empirical protest cycles starts with a set of actors who are assumed to emit protest with some relatively low probability. This simple stochastic or probabilistic model generates some of the waves and spikes seen in empirical data (and also is a plausible representation of the underlying process producing protest events).

The second kind of model seeks to represent a single abstracted process and is more squarely a theory development enterprise. Rather than representing any empirical case, the model represents a unitary process involved in a wide variety of empirical instances, e.g. rational choice, adaptive learning, strategic interaction. Instead of being tied to data, the mathematical model itself is taken as a given because it is a plausible representation of the process of interest and analysis focuses on following the model through to its logical conclusions to predict outcomes. Tests of predictions are generally left to subsequent researchers, who may compare the predictions of competing theories. In our work, we have developed a formal model which assumes that protest cycles are the net result of two diffusion processes in which ideas spread through a population, the first idea being the encouragement to protest and the second being the repression of protest; this model fits the data better than prior simpler models.

In general, models should either represent a unitary process (or the interaction of a few well-studied unitary processes) or should be closely tied to empirical data. Complex models which attempt to represent the interactions among many processes without empirical ties have too many degrees of freedom and are usually impossible to analyze or validate in any systematic way.

3. *Select the basic modeling strategy.* There are many different approaches to constructing mathematical models and different kinds of mathematical representations will be more or less effective in capturing the process of interest. First, there are the number of equations involved in representing the system. Some processes can be modeled with a single equation that can be manipulated by standard mathematical operations and transformed to produce predictions about the outcomes of the process. In some cases, a series of equations representing sub-processes can be resolved into a single equation that predict outcomes. On the other end of the spectrum are models in which there is a separate equation governing the behavior of each individual and these equations interact with each other to produce the outcomes of the system. Single equation or "analytic" models are much more prestigious in the esthetics of mathematics and are advantageous because they allow standard mathematical operations such as integration,

taking derivatives, solving for equilibria or optima, finding asymptotes, solving for thresholds, and so forth. It is much harder (if not impossible) to obtain straightforward "analytic" solutions to multi-equation systems, forcing the analyst to find numeric solutions or generate outcomes via simulation. A single equation can sometimes summarize the behavior of a homogeneous group in which everyone is identical, but multiple equations are generally necessary to represent the interdependent actions of a heterogeneous group. As we work on the problem of modeling protest waves, for example, we are finding that they are best represented as the accumulation of randomly-determined actions by a set of different actors whose actions affect each other. Such models involve creating large arrays in which each row represents an actor and each column represents actors' characteristics (e.g. interest, resources); models which include network ties among actors have another matrix representing the presence or absence of all possible ties. Even if determinate individual decision models are used, such multi-actor systems can be very large and complex, and not amenable to elegant solutions. Instead, the systems are represented in computer programs which perform large numbers of calculations to yield each result. A multi-equation approach which we do not have space to discuss is "cellular automata," in which equations are written to describe how individuals react probabilistically to those near to them, and then these relations predict large-scale phenomena: for example, equations describe how water molecules react to each other, and then can be aggregated into macro phenomena such as river flows. (See Gaylord, D'Andria, and Dandria 1998) Such models may be appropriate for the spread of ideologies through populations.

Multi-equation systems have been used to model the behavior of people in temporary gatherings or crowds. McPhail's cybernetic control theory says collective action is coordinated by individuals adjusting their behavior to bring their perceptual signals in line with a reference signal; models based on this theory show how clusters, arcs, and rings form in crowds as a consequence of common orientations (McPhail 1991, 1993). Feinberg and Johnson (Johnson and Feinberg 1977, 1990; Feinberg and Johnson 1988, 1990a, 1990b) model the well-established empirical phenomenon of "milling" in a crowd, wherein people move around and talk with others near them. Processes built into these models include the influence of other people nearby, the influence of a central agent who is trying to influence the crowd, and the backing away of those who disagree with the emerging consensus. Johnson and Feinberg's insight, that consensus is a product of both influence and exit, is important for understanding the processes of action within a wide range of collectives and the construction and diffusion of a social movement ideology might work according to similar principles.

Another type of modeling strategy specifies the general form of a model based on assumptions about a social process, but the exact shape of the function that represents it is not determined until the model is fit to data. In such cases, there are one or more parameters in the model that are left unspecified until the function is matched to data and the values of the parameters are selected to provide the best fit between model and data. This approach to modeling has been very popular in the diffusion literature and was used in particular to model collective violence diffusion by Pitcher, Hamblin, and Miller's (1978). The core assumption in the model was that the expression of collective violence is controlled by imitation and inhibition processes which are informed by vicarious learning from the outcomes of prior events. For both imitation and inhibition effects, a scale parameter is introduced in the model which relates the relative impact of prior adoptions on later ones.

Although the authors could have treated their model analytically and shown how it

responds to systematic variations in the parameters, they chose instead to estimate the parameters of their model by "fitting" it to empirical data. The approach has its strengths and weaknesses. One weakness is that the model cannot be tested empirically without first using the data to determine the final form of the model. With enough free-floating parameters, models can fit extremely well to nearly any empirical situation. On the other hand, if that model is well-conceived and each parameter has clear substantive meaning, the parameters can be compared to tell something about a social situation that cannot be related by models that do not depend on the empirical data. For example, a parameter conveys the infectiousness of an event within a diffusion process, comparing the estimated parameters across waves that differ in forms of action or occur in different historical or political contexts tell us something about how inter-actor influence responds to these different kinds of conditions.

Further complexity is introduced into the modeling process when we consider the difference between determinate models and stochastic models. Determinate models give a definite single result for each combination of inputs--no matter how many times predictions from the model are computed, the result will always be the same. Rational action models are generally determinate models: at each decision point, the actor is assumed to choose the single action with the highest payoff. A stochastic model is one in which some of the variables are probability distributions rather than single numbers. Adaptive learning models (e.g. Macy 1990) are stochastic models, because at a given time, each actor's behavior is not determined but instead reflects a probability distribution. Our work on modeling protest cycles indicates that stochastic models produce simulated event times series that most resemble empirical protest event time series. Because of the random element involved in stochastic models, the predicted outcomes are considerably more complex--each time the model is run, a different prediction can be produced. These kinds of models, therefore produce probability distributions of predicted outcomes for each combination of inputs, and the character of these distributions become central to understanding the model and assessing how well it models an empirical process.

4. *Start simply and build carefully.* Before developing and testing a model with all the complexity you ultimately wish to capture, start by testing the behavior the simplest possible model for under very simple conditions to be sure that it is free of "glitches." If your model produces the desired pattern, then begin to add factors or features parsimoniously--just enough to capture the process of interest. If the model has sub-processes, validate each completely before allowing them to interact. If you do not fully understand the behavior of simple constituent processes, the results of an elaborate model can be extremely misleading.

Once you begin working with your full model, make certain you test it under the full range of possible conditions. Consider the reasonable range of every variable, and check the behavior of the model under combinations of extremes, e.g., when one variable is zero, another is very large and another is very small. You should also verify that the model is being calculated correctly by running tests of the model with simple "round" numbers and checking the results by hand (or at least through independent computations via a spreadsheet). If there are relevant empirical data or published simulation results from others' work, put those values into the model to see if it generates the correct output.

5. *Face the problem of metric.* Variables in mathematical models inherently are tied to some scale of measurement or metric. Failure to recognize different metrics in a model can distort the results and even make parameters completely nonsensical. Problems with metric

recently made big news when a Mars landing module malfunctioned because data had been entered in the English (inches, feet, miles) system instead of the metric system. Unfortunately, there is almost no discussion of metric in social science, and many published mathematical models fail to treat metric properly. It is all too common to see published models in which parameters have been chosen arbitrarily to give "interesting" numerical results with no attention at all to what those numbers might mean.

There are two ways to handle metric correctly. The first is to explicitly specify the metric for every variable in the model. When all factors have the same metric, the metric can be ignored as it "cancels out." More commonly, attention has to be paid to the translation between metrics. For example, models of the interactions between movements and states (either state policies or state repression) require an explicit attention to metric that is not easily resolved. Mobilization is usually measured as numbers of events, or numbers of participants (although these only incompletely capture the disruptiveness and intensity of mobilization). But in what units should repression or state policies be measured? And what specifically is the relation between a unit of repression and a unit of protest? These are not easy questions to answer, but they must be to produce a meaningful model.

The second choice is to normalize or standardize the model so that every term in the model is either expressed in the same metric or is metric-free. Physical scientists typically use the term "normalize," while social scientists generally use the term "standardize" for the same general concept. Computing standard scores (subtracting the variable's mean and dividing by its standard deviation) is one example of standardization, although not commonly used in modeling. One common strategy is to express some variables as functions of others. Another is to express variables as proportions reflecting their location between meaningful maximum and minimum values. Complex normalizations usually require both strategies. Marwell and Oliver (1993, pp. 27-28), for example, standardize their model by assuming there is a maximum or high provision level that can be set equal to 1 (so that all other provision levels are expressed as proportions of this level), and carefully define contributions, costs, and benefits in terms of these standardized provision levels.

6. *Explicitly identify the scope conditions and assumptions.* All theories, whether verbal or mathematical, positivist or constructionist, contain a variety of assumptions--suppositions taken to be true without proof. Unfortunately, these assumptions are often unacknowledged in much verbal theorizing. One advantage of formal mathematical theory is that the logic of the mathematics itself forces theorists to specify their assumptions, or at least makes them manifest the mathematics of the model. Three kinds of assumptions are important in mathematical models. The first is that the mathematical form used is an adequate representation of a process of interest. In empirical modeling, this assumption need not be taken as a given, but may be based on empirical data. In theoretical modeling, the justification for form is grounded in a belief that it imitates the process of interest. For example, rational action models express in an equation a conscious thought process of weighing cost and benefits that people are believed to use in decision making. In other cases, theorists have written models in which they do not know how exactly the process works, but instead they suppose the process works a particular way and determine the consequences if their assumption were true.

Writing a mathematical model therefore forces you to pin down just how you think things work. Is the relation between protest and repression linear or nonlinear? What kind of nonlinear? Does the relation interact with other factors? Exactly how? As soon as you start

constructing equations or writing computer code, you are forced to become very specific about how you think the process works. This is much harder than just writing a verbal theory that one thing "affects" another. Even so, some assumptions are not always obvious and the theorist must take care to make them explicit. For example, Oliver and Marwell's models permit "rich" people with high resources to make partial contributions, while Macy, Heckathorn, and Kim and Bearman's models assume that actors must contribute all their resources if they contribute anything; these embedded assumptions can make big differences in the outcomes of the models.

The second kind of assumption is called a scope condition. Scope conditions limit the context in which the theory is expected to operate. This principle is all too often ignored in verbal theorizing. For example, consider the classic resource mobilization claim that there are always enough grievances and the real predictor of protest is resources. Not surprisingly, empirical researchers quickly demonstrated that aggrieved people protest more than non-aggrieved people, a proposition that resource mobilization theorists would never have disputed in the first place. And, in fact, the initial statement would have more fruitfully been initially stated as: "for those issues about which there is a grievance, the resources of the aggrieved groups determine which ones will be acted upon," where the underlined clause is a scope condition. If a scope condition is not true, the theory does not apply. Finding examples of empirical instances which do not meet the scope conditions of a theory in no way disproves it. Instead, the concern is whether there are some instances that do meet the assumptions, and whether the model's predictions are true when the assumptions are met.

The third important kind of assumption in mathematical theorizing is a simplifying assumption. In this case, the theorist knows that the process is actually more complex than the model, but some factors are purposely ignored, or relationships are represented with approximations that are known not to be strictly correct. Simplifying assumptions are made so that a model can be made tractable, that is, capable of being analyzed mathematically. All models (and all theories) require simplifying assumptions. For one, they have to ignore some factors that might influence the outcomes. While in reality everything may have some connection to everything else, it is impossible to develop any kind of theory by considering everything at once. Theorists must use boundaries and assume that the factors outside the boundary have insignificant effects. Apart from bounding the model, other simplifying assumptions are often necessary, especially in the initial stages of development. If a simplified model is shown to transcend the simplifying restrictions, then the model is said to be "robust." If, however, the results change dramatically when simplifying assumptions are relaxed, the assumptions must either become scope conditions for the model, or must be systematically varied and analyzed.

7. Analyze your model. Determine its behavior under limiting conditions. Identify the reduced forms of equations. Conduct experiments and map the response surface. Once your model is known to be working correctly, you need to analyze it. It is not enough to tweak the parameters so that you can make it behave like one example or make it generate several different "interesting" patterns. Instead you need systematically to determine how the model behaves with all the possible combinations of inputs and parameters. The key ideas for doing this are the *response surface* and *experimental design*. Although the term is rarely used by sociologists, the concept of a response surface is common in statistics and engineering, and is straightforward. You have an output or criterion variable whose behavior you are interested in, such as the

amount of protest. This is just the dependent variable or "y" that sociologists are used to. If you have only one independent variable or input ("x"), you can plot a standard two-dimensional graph of a line showing how y changes with x. That is a two-dimensional response surface. If there are two independent variables or inputs, you can plot the outcome or y variable as a function of the inputs on a three-dimensional graph, in which case the plot would be a surface, rather than a line. The general concept of the response surface simply extends this idea into n-dimensional space. The goal of response surface analysis is to understand how the output variable changes as a function of multiple inputs.

What can be difficult is to recognize that we are often interested in treating parameters as variables. To take a simple example, consider the linear equation $Y = a + bX$. If a and b are given numbers (such as 3 and 5), the equation $Y = 3 + 5X$ yields a specific line relating Y and X. But in response surface analysis, we could be interested in how the relation between Y and X varies as a and b vary, so we would imagine a 4-dimensional space in which the location of the line relating X and Y moves up and down the Y axis depending on a, and the slope of the line gets steeper or flatter and tips to the right or the left depending on b.

We don't do a response surface analysis of a linear equation because we know that it is exactly the same for all possible values of a and b. However, most collective action models are nonlinear systems in which the shape or form of the function changes with different combinations of the parameters. It is common that the response surface is qualitatively different in different regions of the input space, that there are steep changes or discontinuities between regions, and that maximizing outputs involves optimizing rather than maximizing inputs. In collective action models, there are often thresholds for combinations of inputs below which the output is constant, and above which the output changes with the inputs, often discontinuously. Once you understand the idea of a response surface, it is a straightforward extension to recognize that the output does not necessarily have to be a continuous quantitative number, but can be qualitatively different states. You may also be interested in multiple outputs.

If possible, begin by analyzing your model symbolically using standard mathematical approaches such as algebra and differential or integral calculus to solve for maxima and minima, thresholds, limits, equilibrium states, reduced forms, and the like. Obviously, the more mathematics you know, the more likely you are to be able to conduct these analyses, and it is always worthwhile to spend some time with appropriate texts learning or reviewing basic mathematical approaches to the class of equations you are working with. Even for complex multi-equation models, you can often find analytic solutions for some variables when others are held constant.

Experimental design is another powerful framework for analyzing complex response surfaces, and it is essential that you understand the basic principles of experimental design. In an experiment, you hold some factors constant and systematically vary others. Your research purpose or theory should tell you which elements of your model should be held entirely constant in your analysis. These constants become scope conditions for the results of your analysis.

Your analysis should involve a combination of exploration and focused comparison. First you explore the model, looking for thresholds, limits, equilibria and the like, guided as much as possible by your prior analysis. Be sure to examine the model under the full range of extremes, including unrealistic ones. It is easy to make false assumptions about the realistic ranges of variables, or falsely to extrapolate from too narrow a range of values. Oliver (1993) found that one of the results Heckathorn (1988) reported about group size and social control

changed for very low probabilities of detection of deviance. What extremes are realistic? Try to consider the kinds of empirical situations to which they might apply and come up with some estimates. For example, Oliver developed a model involving the degree of group heterogeneity, operationalized as the standard deviation of a standardized lognormal distribution (mean=1) of contribution levels. Initial analyses were conducted with standard deviations in the .3 to 3 range. A little research revealed that (as of 1990) the US income distribution was roughly lognormal with a standard deviation of .8 of its mean, indicating that this initial range was plausible. However, some additional computation revealed that a standard deviation of 20 was well within the limits of empirical plausibility (e.g. a distribution in which the vast majority give almost nothing and 1 person in 5000 gives \$300 or more, or in which the average time contribution is an hour a year but one in 5000 gives six hours a week). It turned out that the response surface of the model changed dramatically for such high standard deviations. Thus we presented results for standard deviations of .3, 1, 3, 10, and 20 (see Marwell and Oliver 1993, pp. 134-135).

Group size often matters in collective action models. Groups of size two and three have their own peculiar dynamics. Real social groups often involve at least thousands of members, possibly millions. Can a model with, say, 100 or even just 10 actors in it be trusted to give results that would apply to 1,000,000 actors? Different theories would give different answers. What we can say is that you should give some theoretical attention to the group size question, not use very small groups unless you are specifically interested in small groups, and probably run your model with at least two and preferably at least three different group sizes, so that you can get some information about whether group size affects the results. If you are running a model in which group size is a parameter (N) instead of the number of interacting individuals in a model, test group sizes of different magnitudes (e.g. 10, 100, 1000, 10,000, 100,000) that are reasonable for the kinds of groups you are modeling.

After sufficient exploration that you understand your models, design focused comparisons to test theoretical propositions, to compare your model to other models, or to assess the possible impact of some particular change. Determine which elements should be held constant for this purpose, and which should be varied. Design a systematic data collection procedure. One approach is to select values at equal intervals from the full range for each variable. If you have two inputs that are each sampled at six levels in a fully-crossed experimental design including all possible combinations, you would need to generate 6^2 or 36 "cases," while for five inputs that are each sampled at six levels, a fully-crossed design would yield 6^5 or 7776 distinct data points, so sample sizes can increase rapidly as you systematically test a model. If your model is determinate, you only need to run it once for each design cell, and if it can be calculated quickly on a computer, it may not be difficult to program a loop that generates all the input combinations and calculates and stores all the output combinations in a data base. Another possibility is to use a random number generator to select a value for each input from its range, and repeat the experiment as many times as you can afford to. The result of either procedure will be a data set of inputs and outputs that you can plot in various ways. However, whenever the relationships in your model are not linear, equal-spaced intervals will not necessarily give the best information about your response surface. Your initial analysis may reveal that there are particular regions of the space that require more detailed analysis. Because we are unable to convey n-dimensional spaces on two-dimensional paper, presentation of your results usually requires generating graphical cross-sections of different regions of the response

space. You generate a cross-section by holding all but one or two inputs constant at a particular level and generating a two- or three-dimensional graph of the output as a function of the varying inputs.

Even a complex determinate model might take too long to calculate to make it feasible to generate outputs adequately to represent the whole response surface. If the model is stochastic, you need multiple cases per design cell so that you can identify the probability distribution of outcomes in that cell. How many cases per cell? The answer lies in sampling theory, and depends both upon how variable the outcome is for a given combination of input parameters and the degree of precision you desire for your estimate. But even if you are willing to settle for small samples per design cell (and 10 is a very small sample by any criterion), if you multiply that by the large number of design cells you would have to fully represent a model with many inputs, it is obvious that the number of data points involved would quickly become enormous. And, since stochastic models by their nature require much more computation than determinate models, the amount of time necessary to generate those data points could become impossibly large. It is at this point that a solid understanding of experimental design and the principles of mathematical analysis can guide you in designing a smaller more manageable experiment that will generate useful results by helping you to focus on "interesting" regions of the response surface and generate focused comparisons between different regions. If practical constraints force you to hold many potentially-variable factors of your model constant, you should either report these as scope conditions for your results, or be able to explain why these constant factors would not change the dynamics of the variables you are modeling.

Random number generators play an important role in analyzing mathematical models and simulations. Many statistical packages, mathematical programs, and spreadsheets have built-in procedures for generating random numbers from a number of distributions, and some also have built-in procedures or add-in procedures available to assist with simulation procedures. You need to give serious attention to the particular statistical distribution you use in generating a random number, depending on its role in the model. If you are generating random numbers for inputs to use in collecting data on a response surface, you would use a uniform distribution if you want to get equal coverage of all regions of the surface as the random analogue to equally-spaced intervals. But if you know that the output changes more quickly in some regions than others, a more efficient design would use a random distribution that generates more data points near the critical area and fewer where the output changes more slowly. When you are analyzing a determinate model, the choice of distribution just affects the efficiency of data collection. But if there is a stochastic component to your model, the form of the distribution from which probabilities or other random elements is drawn can make a big difference in the outcomes, and you should not blindly use a uniform or normal distribution without considering theoretically what kind of process underlies the random element. A serious discussion of statistical distributions is beyond the scope of this chapter. (An advanced treatment of issues of experimental design and response surface analysis may be found in Myers and Montgomery 1995; see Law and Kelton 2000 for an advanced treatment of principles of simulation modeling and experimental analysis.)

8. *Assess the fit of your model to criterion data.* If the purpose of a model is to represent an empirical instance, testing the fit of the model is obviously required. But even if the purpose is to develop theory, it is appropriate to determine if the theory seems to fit appropriate empirical instances and there is usually at least some data relevant to elements of the model. Fit

can be assessed with respect to the assumptions and the predicted outcomes of a model. The assumption that people seek to maximize their own payoffs can be assessed in bargaining experiments by asking players what factors they are considering (e.g. Michener and Myers 1998a; b). Opinion data can be used to assess whether people's interests remain relatively fixed, as Marwell and Oliver (1993) assumed, or change over time, as Kim and Bearman (1997) assumed, and so forth. Finding some cases that do not fit a theory's assumptions do not disprove it, but if no cases fit the assumptions, the usefulness of a theory is called into question.

Finding data relevant to a theoretical model can sometimes be difficult, but often some indirect evidence can be used. For example, early resource mobilization theory argued that external resources cause protest. But McAdam (1982) presented data on the civil rights movement showing that rises in mobilization preceded rises in external funding, thus sparking a reconsideration of the role of external resources in protest. Models of protest mobilization should be compared to the basic patterns of protest data: wavelike, spiky, and having waves within waves. Theoretical modeling is often useful precisely when relevant data are not readily accessible, but models should still be subjected to basic "reality checks" against what data are out there.

<<FIGURE 1 ABOUT HERE >>

Even when appropriate empirical data exist, "fitting" the model is not necessarily a straightforward process. Among sociologists trained in regression approaches, our first inclination is to calculate a χ^2 or R^2 test. However, in some cases, these tests can be worse than useless. If the data are cumulative event counts, for example, R^2 tests are quite misleading. Consider the plots in figure 1 of the density and cumulative distributions for constant, exponential growth, and diffusion models over time. The correlation between the cumulative logistic diffusion curve and a line with the same minimum and maximum yields an R^2 of .94, even though the correlation between the density functions of the same curves is 0. The correlation between a cumulative logistic diffusion curve and a cumulative exponential growth curve is +.78, yielding an R^2 of .6, even though the correlation between the density functions is actually negative, -.47. Pitcher, Hamblin, and Miller (1978) fit their diffusion model against 25 data series and found R^2 value generally exceeding .95. As impressive as this seems, the results are misleading because the data are cumulative event counts, and any S-shaped function will necessarily have very high correlations with the data. In short, models that produce the wrong basic shape should be readily discarded, but if the basic shape is correct, more rigorous and detailed assessments of fit are required.

9. *Write about your model.* The final step in any research project is communicating the results. Your goal is to contribute to sociology, not just to play with numbers and graphs. Clearly identifying the process(es) you are attempting to model and the scope conditions of the model, and tie these to ongoing issues in the relevant theoretical and empirical literatures. Report on its limits and response surface. Report on its fit (if any) to empirical data. Give a clear account of the theoretical and empirical implications of your model. As much as possible, create a narrative line that explains the logic of the model, its assumptions, its results and its significance that can be followed even by a reader who is unable or unwilling to follow the mathematical reasoning. This is easier to do if you have not cheated on step 1, and have actually read and thought about the broader literature to which your model can speak.

Strengths and weaknesses

As we have indicated, the great strength of mathematical modeling and simulation approaches is the ability to express relationships in clear, unambiguous ways and derive previously-unrecognized results or predictions from them. These approaches are useful for developing abstract theory which may be applied across many instances, and are appropriately compared to abstract verbal theory approaches, rather than to empirical research. Any theoretical representation is necessarily an abstraction that, by its nature, lacks content about particular historical events, ideologies, or personalities. The relationship between theory and empirical data is always a dialectic between the general and the particular. There are often proposed relationships from verbal theory that cannot be readily or clearly stated in mathematical terms, but this most commonly points to ambiguity in the verbal theory, or to the modeler's lack of knowledge of mathematical forms to represent particular kinds of relationships. Mathematical models in sociology are generally much more simple than complex empirical relationships, but so is any verbal theory. Physical and biological scientists building upon a more substantial base of well-confirmed simple relationships have, in fact, been able to build complex representations of complex empirical systems. Obviously, mathematical approaches need to be used in combination with other methods which provide either quantitative or qualitative information about empirical patterns and relationships. Modeling should be viewed as an important complement to empirical research, not a substitute for it. However, the researcher doing modeling will often draw on or inspire other scholars' empirical research, rather than combining modeling and empirical research in the same article or book. Knowledge accumulates when scholars build on each other's work, rather than expecting that any one research project can provide the one perfect definitive answer.

Conclusion

Both mathematical sociologists and postmodernists have been accused of trying to write in ways that others cannot understand, and are often assumed by others to have nothing useful to say if it cannot be said in "plain English." We are not competent to comment on postmodernist writing, but we are prepared to say that people "speak mathematics" not as an arcane jargon to shut out others, but because it is an elegant and effective mode for communicating certain kinds of ideas. We do not believe that formal mathematical models are the only way to do sociology. Rather, we believe that sociologists are bound in a Durkheimian organic solidarity based on a division of labor in which different kinds of theory and methods each play an important role. In this respect, formal models do a particular kind of work. Most importantly, the process of writing a formal model forces you to pin down exactly what you mean, operationalize relationships, and specify mechanisms. Turning thoughts into equations reveals ambiguities and contradictions very quickly. In addition, a good model can permit a "what if" analysis, allowing you to explore possibilities that do not actually exist. If you can adequately represent the mechanisms in "what is," you can explore what would happen if some elements of the system change.

We have mostly written as if the audience is made up of potential modelers, and we hope some of you will be motivated to try it. But we know that most readers are more interested in evaluating others' models than in writing their own, so we end by summarizing the principles to use in evaluating a formal model.

First, what is it about? What empirical phenomena does it attempt to represent? What

examples do the authors give? Do they cite literature or give other basis for suggesting that their image of the empirical phenomenon is correct? Or, if it is a purely theoretical article, what is the core process in the model? Is this a kind of "normative" theory, in which the point is not to model actual behavior, but to provide a baseline by showing what the results would be if certain assumptions were true?

Secondly, what are the scope conditions and simplifying assumptions? These should be spelled out. Some of the scope conditions will be implicit in the way the equations are written, and may be difficult to find if you cannot read the equations. Remember, however, that you should not be focusing on whether you can think of any exceptions to the scope conditions – there will always be cases that do not fit the scope of a theory. Rather you should consider whether there are cases that do fit the scope conditions. Also look for the author's discussions of how the model changes if some of the simplifying assumptions are relaxed: is the model robust?

Thirdly, how is the model analyzed? It is very helpful to the non-mathematical reader to have a narrative which explains how the model works, but the analysis should be more than just story-telling, more than writing equations which can do the same thing as you can do verbally. There should be some analysis that shows how the model changes as the parameters of the model change. There should be a presentation of controlled comparisons within the model, and if it is compared to other similar models, there should be controlled comparisons with the other theories.

Finally, consider the empirical credibility of the model. It is usually not possible to published an extended model and an extended data analysis in the same article. But any presentation of relevant data is a plus, and there should at least be some discussion of the kinds of examples the model should apply to. Look at the outcomes the model produces. Do they seem to fit what you know about the empirical data? Do they seem to illuminate the mechanisms involved? Again, do not reject a model because you know of some counter-example or, worse, because you do not like the implications of the results. And do not accept it just because you can think of one example that seems to fit or you do like the implications. But it is appropriate to bring to bear what knowledge you do have about the phenomenon in evaluating the model.

As one element of a diverse repertoire of methodologies designed to illuminate different aspects of collective action and social movements, mathematical models can make important contributions. We invite you to explore the works we have listed in the bibliography and to undertake some modeling of your own.

References (with some annotations)

- Axelrod, Robert M. 1984. *The Evolution of Cooperation*. New York: Basic Books.
- Coleman, James Samuel. 1964. *Introduction to Mathematical Sociology* New York: Free Press of Glencoe. A widely-read classic.
- Chong, Dennis 1991. *Collective Action and the Civil Rights Movement*. Chicago: The University of Chicago Press. A supply-demand model; careful mathematical analysis; weaknesses are inadequate treatment of metric and poor fit to empirical patterns.
- Cornes, Richard and Todd Sandler. 1996. *The Theory of Externalities, Public Goods, and Club Goods*. Second Edition. Cambridge, New York and Sydney: Cambridge University Press.
- Cortazar, Rene. 1997. "Non-Redundant Groups, the Assurance Game and the Origins of Collective Action." *Public Choice* 92:41-53.
- Davis, Morton D. 1983. *Game Theory: A Nontechnical Introduction*. New York: Basic Books.
- Fararo, Thomas J. 1978. *Mathematical Sociology: An Introduction to Fundamentals*. Huntington, N.Y.: R. E. Krieger. A useful classic.
- Feinberg, William E. and Norris R. Johnson. 1988. "Outside Agitators" and Crowds: Results from a Computer Simulation Model. *Social Forces* 67: 398-423.
- Feinberg, William E. and Norris R. Johnson. 1990a. Elementary Social Structure and the Resolution of Ambiguity: Some Results from a Computer Simulation Model. *Sociological Focus* 23: 315-331.
- Feinberg, William E. and Norris R. Johnson. 1990b. Radical Leaders, Moderate Followers: Effects of Alternative Strategies on Achieving Consensus for Action in Simulated Crowds. *The Journal of Mathematical Sociology* 15: 91-115.
- Gaylord, Richard J. Lou D'Andria, Louis J. Dandria. 1998. *Simulating Society: A Mathematica Toolkit for Modeling Socioeconomic Behavior*. New York: Springer. Cellular automata models.
- Gilbert, G. Nigel, and Klaus G. Troitzsch. 1999. *Simulation for the Social Scientist*. Buckingham ; Philadelphia, Pa.: Open University Press. A general discussion with programs in a programming language called lisp.
- Gould, Roger V. 1993. "Collective Action and Network Structure." *American Sociological Review* 58:182-196.
- Hamburger, Henry. 1979. *Games as Models of Social Phenomena*. San Francisco: W. H. Freeman and Co.

Jones, Anthony J. 1979. *Game Theory : Mathematical Models of Conflict*. New York: Halsted Press.

Hardin, Russel 1971. Collective Action as an Agreeable N-prisoner's Dilemma. *Behavioral Science* 16: 472-481.

Hardin, Russel 1982. *Collective Action*. Johns Hopkins University Press.

Heckathorn, Douglas D. 1988. "Collective Sanctions and the Creation of Prisoners-Dilemma Norms." *American Journal of Sociology* 94:535-562.

Heckathorn, Douglas D. 1990. "Collective Sanctions and Compliance Norms - a Formal Theory of Group-Mediated Social-Control." *American Sociological Review* 55:366-384.

Heckathorn, Douglas D. 1993. "Collective Action and Group Heterogeneity - Voluntary Provision Versus Selective Incentives." *American Sociological Review* 58:329-350.

Heckathorn, Douglas D. 1996. "The Dynamics and Dilemmas of Collective Action." *American Sociological Review* 61:250-277.

Johnson, Norris R. and William E. Feinberg. 1977. A Computer Simulation of the Emergence of Consensus in Crowds. *American Sociological Review* 42: 505-521.

Johnson, Norris R. and William E. Feinberg. 1990. Ambiguity and Crowds: Results from a Computer Simulation Model. *Research in Social Movements, Conflicts and Change* 12: 35-66.

Jones, Anthony J. 1979. *Game Theory : Mathematical Models of Conflict*. New York : Halsted Press.

Kim, HyoJoung, and Peter S. Bearman. 1997. "The Structure and Dynamics of Movement Participation." *American Sociological Review* 62:70-93.

Lave, Charles A. and James G. March. 1975. *An Introduction to Models in the Social Science*. New York: Harper & Row. Widely-read and very useful.

Law, Averill M. and W. David Kelton. 2000. *Simulation Modeling and Analysis* (Third Edition) McGraw-Hill, 2000. A standard simulation text widely used by professionals; the focus is experimental design and using simulation to understand an empirical problem, not programming.

Macy, Michael W. 1990. "Learning-Theory and the Logic of Critical Mass." *American Sociological Review* 55:809-826.

Macy, Michael W. 1991a. "Chains of Cooperation - Threshold Effects in Collective Action."

American Sociological Review 56:730-747.

Macy, Michael W. 1991b. "Learning to Cooperate - Stochastic and Tacit Collusion in Social-Exchange." *American Journal of Sociology* 97:808-843.

Macy, Michael W. 1993. "Backward-Looking Social-Control." *American Sociological Review* 58:819-836.

Macy, Michael W. 1997. "Identity, Interest and Emergent Rationality - an Evolutionary Synthesis." *Rationality and Society* 9:427-448.

Marquette, Jesse F. 1981. A Logistic Diffusion Model of Political Mobilization. *Political Behavior* 3:7-30.

Marwell, Gerald, and Pamela E. Oliver. 1993. *The Critical Mass in Collective Action: A Micro-Social Theory*: Cambridge University Press.

McAdam, Doug. 1982. *Political Process and the Development of Black Insurgency, 1930-1970*. Chicago: University of Chicago Press.

McPhail, Clark. 1991. *The Myth of the Madding Crowd*. New York: Walter DeGruyter.

Mc Phail, Clark 1993. From Clusters to Arcs and Rings: Elementary Forms of Sociation in Temporary Gatherings. In S. E. Cahill and L. H. Lofland, eds., *The Community of the Streets*. JAI Press.

Mc Phail, Clark, W. T. Powers, and C. W. Tucker. 1992. Simulating Individual and Collective Action in Temporary Gatherings. *Social Science Computer Review* 10: 1-28.

Michener, H. Andrew and Daniel J. Myers. 1998a. "Probabilistic Coalition Structure Theories: An Empirical Comparison in 4-Person Superadditive Sidepayment Games." *Journal of Conflict Resolution* 42(6): 830-860.

Michener, H. Andrew and Daniel J. Myers. 1998b. "A Test of Probabilistic Coalition Structure Theories in 3-Person Sidepayment Games." *Theory and Decision* 45(1):37-82.

Myers, Raymond H. and Douglas C. Montgomery. 1995. *Response Surface Methodology*. Wiley. A standard text in its third edition which integrates experimental design and problems of optimization.

Oliver, Pamela E. 1980. Rewards and Punishments as Selective Incentives for Collective Action: Theoretical Investigations. *American Journal of Sociology* 85: 1356-75.

Oliver, Pamela E. 1993. "Formal Models of Collective Action." *Annual Review of Sociology* 19:271-300.

Oliver, Pamela E., and Gerald Marwell. 1988. "The Paradox of Group-Size in Collective Action - a Theory of the Critical Mass .2." *American Sociological Review* 53:1-8.

Oliver, Pamela E., Gerald Marwell, and Ruy A. Teixeira. 1985. "A Theory of the Critical Mass .1. Interdependence, Group Heterogeneity, and the Production of Collective Action." *American Journal of Sociology* 91:522-556.

Olson, Mancur The Logic of Collective Action 1965. The Logic of Collective Action. Harvard University Press.

Owen, Guillermo. 1995. *Game Theory*. San Diego: Academic Press.

Pitcher, Brian L., Robert L. Hamblin, and Jerry L. L. Miller. 1978 the Diffusion of Collective Violence. *American Sociological Review* 43: 23-35.

Rapoport, An. 1970. *N-Person Game Theory: Concepts and Applications*. Ann Arbor: University of Michigan.

Rapoport, Anatol. 1970. *N-person Game Theory: Concepts and Applications*. Ann Arbor: University of Michigan.

Runge, C. Ford 1984. Institutions and the free rider: the assurance problem in collective action. *The Journal of Politics* 46: 154-81.

Sandler, Todd. *Collective action: Theory and applications* Ann Arbor : University of Michigan Press, c1992.

Tools

Many people who write formal models also write their own computer programs to analyze them, using whatever computer language they happen to know. It is possible to do a great deal of analysis using a spreadsheet. There are also more specialized computer programs particularly suited for working with mathematical models or simulations. We list some we have found to give you an idea of what is available. Prices are as of the date we searched; editions and prices are continually evolving. Academic prices for proven members of educational institutions are substantially lower than commercial prices.

1) Mathematics programs. *Mathematica* by Wolfram, Inc. and *Maple* by Waterloo Maple Inc. are complex and very powerful programs which can do symbolic mathematics as well as numerical computations, and can create impressive 3-D graphical images. You can do formal analysis of models with these packages, as well as design simulations. There are ongoing debates between users of each program about which is better, and they are definitely different from each other in their syntax and underlying programming logic. Both are relatively difficult to learn to use. *Mathematica* has a richer library of add-ons and third party software products and books supports a wide variety of modeling and simulation activities. *Maple* is generally less expensive. Academic editions are available for under \$1000, and student editions are considerably less expensive.

Mathematica

Wolfram, Inc.

www.wolfram.com

Corporate Headquarters

Wolfram Research, Inc.

100 Trade Center Drive

Champaign, IL 61820-7237 USA

Sales and order inquiries: 1-800-WOLFRAM (965-3726) or

1-800-441-MATH (6284) (U.S. and Canada only)

Fax: 217-398-0747

Sales: info@wolfram.com

Europe

Wolfram Research Europe Ltd.

10 Blenheim Office Park

Lower Road

Long Hanborough

Oxfordshire OX8 8LN UNITED KINGDOM

Telephone Number: +44-(0)1993-883400

Fax: +44-(0)1993-883800

Sales: info@wolfram.co.uk

Asia

Wolfram Research Asia Ltd.

Oak Ochanomizu Building 5F

3-8 Ogawa-machi Kanda

Chiyoda-ku, Tokyo 101-0052 JAPAN

Telephone Number: +81-(0)3-3518-2880

Fax: +81-(0)3-3518-2877
Sales: info@wolfram.co.jp

Maple.

Waterloo Maple Inc.
57 Erb Street West
Waterloo, Ontario
Canada N2L 6C2
Phone: +1 (519) 747-2373
Fax: +1 (519) 747-5284
Sales (North America): 1 (800) 267-6583
Email: info@maplesoft.com
<http://www.maplesoft.com/>

2) Graphical simulation programs make it easy to construct certain kinds of models with icons and links between icons. Most of these are specialized products oriented to engineers designing manufacturing or business processes, computer networks, and the like. We have found two packages that are suitable for the more general needs of academic social scientists. As compared with programming languages, these programs are much easier to use: it is possible to produce a basic model and get results very quickly. This ease of use makes it possible to focus on thinking about the design of the model instead of figuring out how to write a program and deal with error messages. However, the graphical programs are less flexible and powerful. Whether the limitations are a problem depends upon your particular model.

Stella from High Performance Systems, Inc. is a general purpose simulation package with a graphical interface which is especially well-suited for representing feedback processes that occur over time, e.g. predator-prey relations. You represent the model as stocks and flows, and can specify the exact equation representing a relationship. Unlike most graphical packages, *Stella* produces an equations page to complement the graphical representation. It has special features for setting up computer- or web-based interfaces for educational presentations. The user interface is very easy to use although it can be cumbersome for large batch-oriented data generation, but it can link to spreadsheets for large input-output tasks. There are stochastic functions. The more expensive research edition can handle arrays. The documentation emphasizes the principles of model-building. There is a less expensive basic edition, and a research edition which has much higher capacity and extra features. Student pricing starts under \$100, educator pricing is \$300-\$550 depending on version. A free demo is available for download. (Note, the same firm produces *iThink* which is exactly the same program marketed to business applications.)

High Performance Systems, Inc.
45 Lyme Road, Suite 300
Hanover, NH 03755-1221
Phone: (800) 332 1202 (603) 643 9636
<http://www.hps-inc.com/>

Extend from Imagine That, Inc. is a general-purpose simulation package with pedagogic

materials. Everything is an icon, even adding and subtracting, and the icons are highly specialized, not really equation or functionally oriented. Icons generate C-like code which apparently can be edited, although that feature was not available in the demo we examined. Models can generate processes which can be animated with icons you choose. You can do sensitivity analyses, and interface with other programs. Academic licenses begin at \$350, pricing a limited edition for students or evaluation purposes begins at \$60. You can download a free demonstration version.

Imagine That, Inc.

6830 Via Del Oro, Suite 230

San Jose, CA 95119 USA

408-365-0305, fax 408-629-1251

email extend@imaginethatinc.com

<http://www.imaginethatinc.com/>

3) Simulation languages are powerful programming languages with special constructs to make it easier to write and test simulations. They are generally easier to learn to program than general programming languages (or *Mathematica/Maple*), but require more learning than graphical packages. The firms offering simulation languages also market graphical packages.

Wolverine Software sells *GPSS/H*, a version of the longstanding GPSS simulation language, and *SLX* which is a powerful multi-layered simulation product. A student version of *SLX* is available free and a student version of *GPSS/H* is available for \$40; these are limited versions which can be used for evaluation purposes. The student version of *GPSS/H* includes a textbook and examples. Academic versions are \$750 and \$1000.

Wolverine Software Corporation

2111 Eisenhower Avenue, Suite 404

Alexandria, VA 22314-4679

(800) 456-5671

(703) 535-6760

Fax: (703) 535-6763

mail@wolverinesoftware.com

<http://www.wolverinesoftware.com/products.htm>

Simscrip 11.5 and *Simprocess* are products of CACI, Inc. *Simscrip* is a free-form, English-like general-purpose programming language with simulation constructs and built-in features for experimental design and response surface analysis. Its core simulation constructs are entities and activities. It is built on a C++ compiler, available for mainframe and PC, and has a menu oriented user interface. It has graphical user interfaces and animation capabilities, self-documenting code, and built-in constructs for discrete-event and combined discrete/continuous process-oriented simulations. CACI also has a graphical-interface product called *Simprocess* which is available in a student version which can be downloaded as a demo. Its orientation is entities (e.g. customer calls) going through activities; it is well designed for cost accounting and simulating business processes, but appears too specialized for more general sociological modeling. Pricing information is not available without talking to a representative.

CACI

Corporate web page

[Www.caci.com](http://www.caci.com)

Simscrip and Simprocess info

<http://www.caciasl.com/>

simscrip@caciasl.com

619-542-5228.

Some other useful resources

Web page on social simulation:

<http://www.soc.surrey.ac.uk/research/simsoc/simsoc.html>

Has lots of links to other social simulation sites. Well-organized, good set of links.

The Journal of Artificial Societies and Social Simulation (JASSS) is the premier journal in the field of social simulation and is the best source for a diversity of social simulations, as well as discussions about simulation principles.

<http://www.soc.surrey.ac.uk/research/simsoc/cress.html>

The *Journal of Mathematical Sociology* publishes a wide variety of mathematical models in sociology, including many relevant to collective action, and is the best single source for exploring the diversity of mathematical modeling in sociology.

Editor:

Patrick Doreian

Department of Sociology, University of Pittsburgh

Pittsburgh, PA 15260, USA

Phone: (412) 648 7537

Fax: (412) 648 2799

Email: pdjms+@pitt.edu

Subscriptions are held by most libraries. Publisher is the Gordon and Breach Publishing Group.

Figure 1. Comparing three models of protests over time: constant (the amount of action is steady across the time period), growth (the amount of action increases by a constant percentage of the prior amount), diffusion (the amount of action is a function of the product of the proportion who have already acted and those who have not acted yet). The top panel shows the density function, the bottom panel shows the cumulative amount of action over time. All three curves begin and end at the same cumulative amounts of action, but get there in different ways. The density function for the constant action model has a zero correlation with the others, while the growth and diffusion models are negatively correlated, but the cumulative functions for all three are highly positively correlated.

