Mathematical Sociology and Empirical Social Research

Kazuo Yamaguchi
The University of Chicago
5801 South Ellis Ave. Chicago, IL 60637 USA

Abstract

This article reviews and discusses various ways to strengthen the linkage between mathematical sociology and empirical research. It distinguishes between the use of mathematics in sociology for theory construction and for empirical data analysis, and discusses how the linkage has been made and how it can be strengthened in each category of use. Concerning the use of mathematics for theory construction, the article compares the integrated division of labor between theory construction and empirical research to traditional divisions of labor, emphasizing the relative importance of the former. Concerning the use of mathematics for data analysis, the paper identifies (1) indexing and measurement, (2) mathematical methods for heuristic analysis of qualitative data, (3) substantively motivated statistical models, and (4) hybrids of mathematical and statistical models, as four forms of linkage between mathematical sociology and empirical research, discussing the accomplishment and prospects for the future for each form.

Key Words and Phrases: theory construction, mathematical methods, mathematical models, statistical models

1. Introduction

While mathematical sociology has been in development for many years, the use of mathematics has never had a central position in sociology as it has in economics. The use of mathematics in the social sciences serves two major functions in social research: theory construction and empirical data analysis. This article is concerned with identifying various ways to strengthen the linkage between mathematical sociology and empirical social research in each of these two functions. I consider this linkage to be critical in order for mathematical sociology to become more central in main-stream sociology than it currently is.

While deductive theory construction based on mathematics can be cumulative, there has been little development of such a cumulative theory in sociology regardless of whether it is mathematical or not. The primary reason as to why deductive theory construction has not played any central role in sociology stems from the fact that a general "grand" sociological theory cannot be based on a formally specifiable set of elementary assumptions as with microeconomic theory. Instead, sociological theories...
are typically middle range theories about specific phenomena, such as those about status attainment, occupational mobility, labor market segmentation, residential racial segregation, gender inequality in wages, gender inequality in the division of household labor, class identification, etc., in the studies of social stratification and inequality. Examples of middle range theories in studies of social psychology including those on rational choice are theories about socialization, reference group, relative deprivation, cognitive dissonance, exchange through network, trust in the prisoner’s dilemma situation, gender-role attitude, racial prejudice, etc.

The contributions of mathematical sociology in the elaboration of some middle range theories have been quite remarkable in my opinion, while they may not have been definitive. Nevertheless, middle range theories are very rarely taught in courses of contemporary sociological theories based on the use of mathematics. Instructors of contemporary sociological theory courses will mention James S. Coleman as the pioneer of rational choice theory in sociology, to represent the utilitarian tradition of sociological theories. However, I doubt that any of his specific theories based on mathematics or some formal expressions such as those about power in exchange relations, norm generation, zealots, etc. (Coleman 1990) will be taught in those courses. We can expect these instructors to refer to Coleman’s theory on social capital (1988), or even describe “the Coleman boat” (Coleman 1986) which is Coleman’s schematized theoretical frame indicating that theories on a macro-macro linkage must be accompanied by explanations for the intervening macro-micro and micro-macro linkages. However, social capital is in fact not among his studies of mathematical sociology, and the Coleman boat is just a theoretical frame of reference as the AGIL paradigm is in the theory of Talcott Parsons (1951).

Occasionally, some courses of contemporary sociological theories, especially if they are taught by instructors who like formal theories, may refer to select studies based at least in part on the use of mathematics because they are classics in mathematical sociology, such as Blau’s theory (1964) on social exchange, Schelling’s theory (1971, 1978) on residential racial segregation, Boudon’s theory (1977[1982]) on relative deprivation, Coleman’s theory (1972, 1990) on collective action, Granovetter’s theories on the strength of weak ties (1971) and on threshold behavior (1978), or Watts’s theory (1999) on small world phenomenon. However, it is very rare for these studies to be described as contributions of mathematical sociology to sociological theory.

A major question, then, is this: why have such contributions of mathematical sociology failed to become significant in sociological theory? Indeed, despite the contribution of mathematical sociology to middle range theories as described above, it has failed to become a foundation for studies outside of mathematical sociology, with a few exceptions such as Granovettor’s theory on the strength of weak ties. This makes a
sharp contrast, for example, with the centrality that Gary Becker's theories have in economics on a wide variety of topics of sociological interest such as crime, social interaction, family, and habit formation (Becker 1970, 1974, 1981, 1996).

It is true that sociology has been a positive science based primarily on inductive theory construction from empirical research, rather than a science based primarily on deductive theory construction for which the use of mathematics is especially helpful. However, this does not seem to be the cause of mathematical sociology's lack of centrality in sociology. Sociological theory has remained inductive not because it is antagonistic to deductive theory construction, but simply because sociologists have failed to develop such theories. When Blau published *Inequality and Heterogeneity* (1977) which employed deductive theory construction, the American Sociological Association welcomed and praised the book by giving it the Sorokin Award. Although Blau did not use mathematics in his book, his theory encouraged its formalization based on the use of mathematics (Skvoretz 1983). However, Blau's book did not become a path-breaker in changing the mode of theory construction in sociology. The reason for its lack of impact on later theoretical studies in sociology is a matter of speculation, but it seems to be partly because Blau's subsequent book with Schwartz (Blau and Schwartz 1984), which provided empirical tests of some major new hypotheses on inter-group relations derived in the Blau book, was not received with as much enthusiasm. The book showed that Blau's predictions held only under various strong qualifications such as socioeconomic status and race playing much more important roles than other individual attributes in determining intergroup relations, but these qualifications did not lie within the original scope of the theory. In other words, while Blau's macro social theory provided elegant and coherent explanations for some basic facts that are already well established before his book, it was not unconditionally supported by empirical research regarding new hypotheses.

Since the reason why contributions that mathematical sociology has made so far have failed to obtain centrality in sociological theory is not due to antagonism toward deductive theory construction, it requires a different explanation. I consider the major or most important reason to be the weakness of relationship between mathematical sociology and empirical research. In particular, the limitations of Blau's theory seems to be also applicable to many contributions that mathematical sociology has made in elaborating middle range theories: while they provided elegant and coherent explanations for facts which had already been known, they did not produce many original and significant new findings that could enhance empirical ground for the theory. For example, Boudon (1977[1982]), Schelling (1971), and Watts (1998) each provided an elegant explanation for relative deprivation, residential racial segregation, and small-world phenomenon, respectively, but the major facts that these theoretical studies
explained had been documented even if the social mechanism of generating these facts had not been known well before these studies. Although Boudon’s theory was elaborated further by Kosaka (1986) and Yamaguchi (1998) and some empirically testable new hypotheses were proposed there, their hypotheses remain largely untested with empirical data. Clark (1991) once tested the Schelling theory on residential racial segregation with empirical data, and Mare and Berk (2004) have recently obtained a National Science Foundation research grant to elaborate and test Schelling’s theory more systematically. However, this shows that although it has already been more than 30 years since Schelling introduced his theory, the empirical basis of the theory has not yet been well-established regarding the underlying mechanism that he posited. Perhaps Watts’s theory on the small-world phenomenon may become an exemplary exception because in his own subsequent research (2003) he tries to strengthen both empirical and theoretical grounds of his theory. Generally speaking, however, empirical follow-ups to contributions that mathematical sociology made towards middle range theories have been very limited, partly because of what I refer to below as the integrated division of labor between mathematical modeling and empirical research and the use of hybrid models has been rare in sociology.

Sometimes the main concern of mathematical sociologists is even opposite in direction: rather than using a mathematical model to derive hypotheses to predict unknown facts that could enrich the empirical basis of the theory, they often spend their energy towards seeking an alternative formal explanation for empirically well-known facts. A typical example is the many attempts to explain why people tend to cooperate, rather than defect, in a one-time prisoner’s dilemma game situation. While this is certainly a central theoretical issue for understanding rationality because the standard theory predicts defection as the rational choice in this situation, it is an issue of little concern for the majority of sociologists. It may appear to outsiders that mathematical sociologists are more interested in demonstrating their cleverness by becoming the scholar to solve a famous puzzle rather than providing solutions for important real-world social problems. While the study of a fundamental issue is certainly important, I believe that mathematical sociologists should spend their energy in such a way that they contribute more towards the enrichment of new empirical findings that will help us understand the real world better.

As I stated at the beginning of this introduction, the second function of mathematics in sociology is its use for empirical data analysis. The use of mathematics for data analytical purposes has been as significant as the use of mathematics for theory construction. Here, I exclude the use of mathematics for statistical theory and the use of mathematics for general-purpose statistical modeling, but include the use of mathematics for statistically modeling specific social phenomena. Hence, what I refer
to here as mathematical sociology include studies which are traditionally classified as methodology rather than as mathematical sociology. In economics for example, discrete choice models and recent structural econometrics reflect some combinations of statistics and theoretical modeling of substantive phenomena. Many indices describing macro-economic characteristics are also based on the use of mathematics.

The use of mathematics in sociology for data analytical purposes rather than for theory construction includes, though not necessarily exclusive, (1) indexing and measurement, (2) mathematical methods for heuristic analysis of qualitative data, (3) substantively motivated statistical models, and (4) hybrids of mathematical and statistical models, the descriptions of which I elaborate below. More generally, I will identify several different approaches to strengthen the linkage between mathematical sociology and empirical social research for both theory construction and data analysis, provide a selective review of the literature, and discuss possibilities for the future. My review is selective because this article is not intended to present a comprehensive review of the literature, but rather to identify various possibilities of enhancing the quality of linkages between mathematical sociology and empirical social research by illustrating these possibilities by selected studies which I am relatively familiar, including my own.

2. Distinct Approaches to Strengthening Ties between Mathematical Sociology and Empirical Social Research

2.1. The Use of Mathematical Models for Theory Construction

2.1.1. Integrated Division of Labor

A representative way of linking or integrating mathematical models and empirical research is to have a division of labor between the two in such a way that the study employs mathematical models to generate theoretical propositions which can derive some empirically testable novel hypotheses, and then uses a statistical analysis of survey data to test those hypotheses. This is a dominant mode of empirical research in micro economic studies: they typically assume some model of rational choice, derive some novel propositions and corollaries mathematically by assuming actors' utility maximization, derive empirically testable hypotheses, and test them with survey or other observational data using some econometric models. However, this type of division of labor is still rare among sociological studies. Exceptions apply to some studies based on rational choice theory because such studies, like microeconomic studies, permit a derivation of new empirically testable hypotheses based on the assumption of utility maximization. Examples include Yamaguchi and Ferguson’s study (1995) on birth stopping and birth spacing, Montgomery’s study (2003) on religious economies, and Youm and Laumann’s study (2003) on the division of household labor. A volume
edited by Blossfeld and Prein (1997) on rational choice theory and large-scale data analysis emphasizes the importance of this approach and includes chapters that provide additional examples of this type of division of labor.

It seems to me that whether rational choice theory will become one of the main-stream theoretical paradigms in sociology depends heavily on support of the theory by analysis of survey data, rather than by an analysis of experimental data or simulation data. We all know that this is not an easy task, because compared with phenomena that economic studies are interested in, phenomena that sociological studies are interested in require modeling of institutional or structural constraints on choices of actions and/or interdependence of people’s choices, and therefore the modeling of social phenomena tends to be much more complicated than that of economic phenomena. However, some economists are also already undertaking the econometric modeling of social interdependence (e.g., Durlauf and Young 2001), and the statistical handling of some unobservable factors that characterize game-theoretic situations such as auctions (e.g., Paarsch and Hong, forthcoming). Unless we also undertake similar methodological elaborations to enable testing of our theoretical models directly with survey and other observational data and reduce our dependence on experiments and simulations to support our theories, the empirical ground of the theories we derive mathematically will remain weak, and those theories will not likely to be influential in main-stream sociology based on solid empirical evidences. I will later discuss the issue of methodological elaboration further.

2.1.2. Traditional Divisions of Labor

What I refer to as traditional divisions of labor also employs mathematical models to represent a theoretical model to derive empirically testable hypotheses, while data analyses to test those hypotheses are separated from this procedure. However, theory construction and empirical research in the traditional divisions of labor are less integrated than the approach described above because a simultaneous test of the theoretically derived hypotheses with survey and other observational data is not provided. The traditional approach has variety within itself, and includes cases such as:

(1) Cases where no direct tests of hypotheses are provided, but the study presents some analyses of empirical data to indirectly examine the consistency of some empirical results with implications from the mathematical model and/or it provides references to the literature that report findings from survey data analyses which are consistent with those implications;

(2) Cases where no analyses of survey and other observational data are provided, but the study provides analyses of experimental data or cite results from others’ experimental
studies to demonstrate the consistency between implications from the mathematical model and the results from analyses of experimental data;

(3) Cases where no data analyses are presented, but the study employs a mathematical model to explain the underlying mechanism of well-known but often theoretically puzzling empirical facts, or to formalize existing theories which are based on well-established empirical facts.

(4) Cases where mathematical models are employed to provide empirically testable hypotheses including novel ones, but their tests with empirical data are left for future studies.

There are relatively few examples of the first case, but Iannaccone’s study (1988) of the church and religious sects, Watts’s study (1999) of small-world phenomenon, and Buskens’s study (2002) of network and trust belong to this category, as well as Kimura’s reconsideration (1989, 2002) for Olson’s proposition (1971) on the negative group-size effect on the attainment of group goal. While these studies show the consistency of some empirical findings with implications from mathematical models, the major emphasis is on the explanation for the mechanism of generating certain outcomes, rather than a rigorous test of the theory with empirical data. This approach is certainly the first positive step toward establishing a linkage between theoretical models and empirical research, although the kind of consistency tests they provide is not as assuring for the theory as a direct test of hypotheses with empirical data.

There are many examples of the second case because many social psychological studies employ experiments, even though only a relative minority among them employs mathematical models. Again, studies which are based on rational choice theory tend to be more formal than others and fit into this category. Examples include Bonacich’s study (1990) on communication dilemma and Lawler and Yoon’s study (1996) on commitment in exchange relations, both of which include original experiments, as well as my study on power in complementary and substitutable exchange relations (Yamaguchi 1996), which refers to experimental results from others’ studies to show the consistency of these results with the model’s predictions. The use of experimental data rather than survey data in this group is mainly due to the nature of mathematical models, which are more difficult to test with survey data than with experimental data. Furthermore, in order to test these models with survey data, the models need to be extended to include (1) a random error term and (2) a parameter to characterize population heterogeneity, as I describe later in the discussion of hybrid models.

The third case is a popular form of the traditional division of labor. The classical studies of residential racial segregation by Schelling (1971, 1978) and of relative deprivation by Boudon (1977[1982]) belong to this category. Other examples in this category include Land’s formalization (1970) for Durkheim’s theory on the social

While many studies in this category are illuminating, they also have serious limitations. The major limitation is that the explanations provided in these studies, while useful in clarifying the underlying mechanism, remain strictly hypothetical because the consistency of a theoretical model’s predictions with known facts does not guarantee by itself that the proposed theoretical explanation holds empirically regarding the underlying process through which the outcomes are generated. They thus require additional empirical supports to be validated, and there are two ways to provide this. One is to conduct empirical research that includes some information about the mechanism itself, and to test the theoretical model in more depth. Alternatively, we may obtain further implications from the theoretical model and derive new propositions that are not yet tested empirically, and confirm the theory by showing that those additional propositions are supported by empirical data.

The fourth case is probably the most numerous, although I have not systematically surveyed mathematical sociology literature in order to confirm this. As a division of labor, this case is incomplete. Sometimes, the omission of empirical evidence for a theory is quite understandable because of the difficulty in doing so. Granovetter’s theory on threshold behavior (1978; Granovetter and Soong 1983) is a case in point: since the collective outcome often depends sensitively on the distribution of thresholds in the population, it is impossible to test the model’s predictions without knowing the threshold distribution, which is difficult to measure empirically. It is necessary to develop research methods to get more directly relevant data than standard survey data to test hypotheses concerning this type of phenomenon. We may also elaborate the theoretical model to obtain some further theoretical implications that extend the scope of empirically testable hypotheses. I have shown (Yamaguchi 1998, 2000b) that threshold behavior appears as a result of population heterogeneity in the costs and benefits of taking a particular action when these costs and benefits depend on the proportion of people in the population who have already taken the action. Such a reformulation of threshold-behavior model as a rational choice model permits us to derive additional empirically-testable hypotheses.

While we can continue our studies in mathematical sociology that are classified as employing the traditional division of labor, we should also try other forms of linkage between mathematical modeling and empirical research, as we already know that the
dominance of this approach has prevented mathematical sociology from exerting influence on main-stream sociology that requires a solid empirical basis.

2.2. The Use of Mathematics for Data Analytical Purposes
2.2.1. Mathematical Indexing

Among various uses of mathematics for data analytical purposes in sociology, indexing has played major roles. Mathematical models can provide useful indices for empirical research. An index can characterize individual positions in collectivity when we assume certain structure of relations among the positions, or individual states from criteria that involve certain mathematical rather than purely statistical reasoning. An index based on a mathematical model can also characterize collectivity regarding its structural characteristics, and also population states based on certain mathematical, as opposed to purely statistical, reasoning.

A classical example is that of inequality indices, such as the Gini index (e.g., Allison 1978) and Theil’s information theoretic index (Theil 1972), compared with a purely statistical index such as the coefficient of variation in income. Indices obtained by standardization, including total fertility rate and the Laspeyres consumer price index, are based on a counterfactual assumption about what the quantity in question would be under a hypothetical “standard” situation, for which a mathematical reasoning would certainly operate.

More sociological representatives are various measures to characterize social networks, such as Bonacich’s measures (1987) of power and centrality, and related centrality measures (Friedkin’s 1991). Friedkin and Johnson’s measure (1990) of influence for a given network of communication is also a generalization of a centrality measure. Coleman’s measure (1972, 1990) of power in exchange relations and Yamaguchi’s extension (1996) of the Coleman measure for complementary and substitutable exchange relations are other examples. These are measures of individual positions when a set of relations among the positions and some other external conditions are given. On the other hand, Jasso’s justice evaluation function (1978, 1990) characterizes individual states based on certain mathematical criteria to evaluate justice. Buskens and Yamaguchi’s measure (1999) of the mean diffusion time for information flow characterizes a directed relation between two positions for a given network of communication. Some other indices characterize collectivity, or the population. Yasuda’s (1964) and Boudon’s (1970) mobility indices, which were influential before the advent of loglinear analysis of mobility, are examples of these indices. Another very important set of indices for characterizing the population are segregation indices for racial/ethnic segregation of residential areas (e.g. Massay and Denton 1993; Reardon and Firebaugh 2002). Indices such as inequality indices and segregation indices are often
compared among themselves regarding their relative merits and demerits by examining
the satisfaction of certain desirable principles for the indices.

Some indices are based on specific theoretical assumptions for their derivation. For
example, Coleman’s and Yamaguchi’s measures of power in exchange relations are
based on the assumption of rational exchange based, respectively, on maximization of
the Cobb-Douglas function, and the CES function. In the case of Friedkin-Johnson’s
measure of social influence and Buskins-Yamaguchi’s measure of mean diffusion time,
the assumption is of the first-order Markov chain. In the case of Yasuda’s mobility
index, the assumption is about the criteria of separating circulation mobility from
structural mobility to measure the openness of a society.

Indexing has been the most influential linkage between mathematical models and
empirical research regarding the use of mathematics for data analysis in sociology.
However, I expect its utility to decrease in the future. A major reason why indices
have been useful is that statistical models for the entire data from which we obtain
indices were not developed. If we can model the entire data with the model’s
parameters characterizing the structural properties of the data, then we will use
information from the parameter estimates rather than information from indices. For
example, when we began to characterize data in mobility tables using parameters of
loglinear and related models and attained the aim of separating circulation mobility from
structural mobility by the use of these models, a mobility index such as Yasuda’s became
obsolete and was replaced by a set of estimates for parameters of the models. Indices
used for measuring gender segregation of jobs are being replaced by use of the
association models that Charles and Grusky (1995) have developed for modeling the
relevant data. Similarly, to the extent to which we are able to model the entire data that
characterizes residential segregation, network structure, or social exchange, many
descriptive indices will be replaced by estimates of the structural parameters of these
models – although the use of some indices will continue because of their theoretical
foundations.

2.2.2. Mathematical Methods for Heuristic Analysis of Qualitative Data

Some of the so-called qualitative methods are based on the use of mathematical
models. A classical case is White’s analysis (1973) of kinship relations based on
Boolean algebra. A more contemporary version is Ragin’s method (1987; Kanomata et
al. 2001) of macro-sociological analysis based on Boolean algebra, and its extension
with the fuzzy-set method (Ragin 2000). Franzosi’s set theoretic analysis (1994) of
narrative data and Abbott’s sequence analysis (Abbott and Hryck 1990; Abbott and
Barman 1997) based on Gibbs sampling, also belong to this category. These are
non-statistical mathematical methods for data analyses. Regarding the method of macro
comparative analysis based on the use of Boolean algebra, Ragin claims that it represents the "case-centered approach" which contrasts with the "variable-centered approach" (implying regression-type statistical analysis), and argues that the former is better than the latter according to certain criteria, including the retention of information about diversity and complexity of the original data (see Goldthorpe 1997 for a related debate). However, I consider a more useful approach than Ragin's approach to be an integration of the two by a stochastic extension of Ragin's method. Block modeling which was introduced by White et al. (1976) was originally algebraic in nature but was extended to be stochastic by Wasserman and Anderson (1987), and this extension has produced more informative results (e.g., Frank and Yasumoto 1998). An extension of the model into a stochastic model permits, above all, (1) a statistical goodness-of-fit test of alternative models with data, and (2) the use of covariates. While a stochastic extension of Ragin's method may be somewhat complicated, it seems to me to be a useful extension because the purely algebraic approach lacks a test of significance about the model's fit with the data and a reliable criterion for choosing among alternative models, and thereby sets a serious limit on the generalizability of findings.

2.2.3. Substantively Motivated Statistical Models.

Some statistical models have been developed from substantive motivations. While they are statistical models, these models are specially designed to analyze specific substantive phenomena rather than to address purely methodological concerns with the objective of developing generally applicable analytical tools for a given form of data. The developments of these models may be regarded by mathematical sociologists as methodological studies rather than studies in mathematical sociology, if mathematical sociology is confined to studies that contribute to theory construction. However, both indexing and mathematical methods for heuristic analysis of qualitative data are often regarded as mathematical sociology while they are also methodology, and I believe that a narrow conception of mathematical sociology is not beneficial towards strengthening the linkage between mathematical sociology and empirical research.

Some studies that belong to this category are:
1. Sobel's diagonal reference model for mobility effects (Sobel 1981, 1985);
2. Strang-Tuma's diffusion process model (1993);
3. Yamaguchi's mover-stayer mixture model for hazard rates for the analysis of "permanent employment" in Japan (1992);
4. P* social network models by Wasserman and Pattison (1996; Pattison and Wasserman 1999);
5. Yamaguchi's dyadic nonindependence model for ego-centric networks (1990);

Sobel's diagonal reference model for assessing mobility effects solved issues of identification of origin effects, destination effects, and mobility effects, and is a classical exemplary case where a statistical method was developed to address a specific substantive issue of measurement. Strang and Tuma have developed a specific hazard rate model for the analysis of the diffusion of an event, such as medical innovation among physicians. Their model separates contagion effects from the effects of "intrinsic tendencies" for the event to occur, and the contagion effects include parameters for the "susceptibility" of "non-infected" subjects, those for the "contagiousness" of "infected" subjects, and those for "proximity" between infected and non-infected subjects, all with covariates to characterize heterogeneity in the values of these parameters. Yamaguchi's mover-stayer mixture model of hazard rate tries to distinguish covariate effects on the ultimate probability of event non-occurrence from covariate effects on event timing, to analyze the determinants of "permanent employment" separated from the determinants of employment duration in Japan. The model was also used to separate covariate effects on birth stopping from those on birth spacing (Yamaguchi and Ferguson 1995). P* models developed by Wasserman and Pattison are logit models for the analysis of social network data, and together with the pseudo-likelihood method for parameter estimation, they can incorporate various aspects of dyadic nonindependence, especially the estimation of triad effects and other structural properties of Markov graphs. Yamaguchi has developed Goodman-type association models (Goodman 1979) for the analysis of ego-centric network data to handle dyadic nonindependence by using parameters to characterize homophily and social distance, not only in the ego's choice of each friend, but also in the form of interdependence among the ego's choices of multiple friends. Charles and Grusky have also developed Goodman-type association models to analyze gender-segregation of jobs.

These are all statistical models, and can be classified as linear regression models (Sobel's models), hazard rate models (Strang-Tuma model and Yamaguchi's mixture model), or loglinear and related association models (P* models for social network data, Yamaguchi's model for ego-centric social network data, and Charles-Grusky's model for the analysis of gender segregation of jobs). However, they all share the aspect that they were all developed for substantive motivations to analyze specific social phenomena. Except for Sobel's studies, these models were all introduced during the 1990s, and show that the development of substantively-motivated statistical models is coming of age in sociological methodology.

2.2.4. Hybrid Models: Theoretically-Oriented Statistical Models
By hybrid models, I refer to models that include the combination of statistical modeling and mathematical reasoning to address substantive sociological issues. Theory construction and empirical data analysis are integrated, but without a division of labor between the two, in these models because statistical models to be tested with empirical data are, at least in part, theoretically derived. I consider that these hybrid models have the greatest potential in linking mathematical sociology to empirical sociological research, although technically it is probably the most difficult type of linkage. Rather than parameterizing certain quantities of substantive interest in statistical models as in the case of substantively motivated statistical models, the entire model is formulated in hybrid models by incorporating certain mathematically-formulated substantive reasoning, and is then extended as a statistical model with parameters of heterogeneity for regression. There are two subgroups in this category.

The first category includes models that assume some behavioral principles such as the type of utility maximization, and in this regard they resemble discrete-choice models in econometrics. However, the following models differ from typical econometric models based on utility maximization because of its incorporation of certain structural factors. Examples in this category include:

1. Coleman-Hao’s regression model of exchange and power (1989);
2. Logan’s two-sided logit model (1996, 1997); and

The Coleman-Hao model introduces a stochastic component into Coleman’s model of collective action such that certain effects of observable resources which actors are assumed to use on exchange outcomes can be estimated with empirical data. Logan’s two-sided logit model assumes a specific matching mechanism between persons and jobs in the labor market, and a utility maximization principle for actors, enabling the estimation of covariate effects on people’s choices. Snijders’s model is concerned with the analysis of network formation and of choice of a tie in the actor’s ego-centric network, and is also based on the principle of utility maximization for a function of “satisfaction” with actors’ network relations to others.

In developing sociological models that have explicit behavioral components, it seems to me that it is important to consider both relaxation and constraint of the standard rational choice assumption, as long as this can be done without losing formality. By relaxation, I refer to an incorporation of certain psychological or social-psychological elements to permit modification of the standard rational choice assumption. By constraint, I refer to an imposition of certain structural constraints on choices, or on alternatives for choices that actors can make. This is desirable not only because there exists strong resistance among many sociologists in accepting the standard rational choice assumption without any constraints, but also because the standard assumption of
rational choice may not reflect the realities of social behavior very well. In addition to various methods to incorporate the constraints that social network imposes on actors’ choices (Bienenstock and Bonacich 1992; Buskens 2002; Chwe 1999: Raub and Weesie 1990; Sato 2002; Yamaguchi 1998, 2000a), we already have a variety of formal methods to incorporate other elements including, but not limited to, roles (Montgomery 1998, 2000), justice (Jasso 1990), framing of choices (Kahneman and Tversky 1979), reputation (Raub and Weesie 1990), risk aversion in a trust-game situation (Raub and Snijders 1997), uncertainty aversion and personal-efficacy bias in binary choice (Yamaguchi 2000b), and socialization as a rational commitment to a strategy (Yamaguchi 1998), to list some examples. However, an incorporation of interdependence of actions and/or structural constraints on choices certainly makes models complicated, and extending the mathematical model to a statistical model by making it stochastic, and introducing a parameter for heterogeneity for covariate effects to test the extended model with survey data are additional challenges. Furthermore, even when a model testable with survey data is developed, the issue of its parameter estimation may become another hurdle due to the issue of nonlinearity. Hence, we are facing very demanding tasks. However, further development and use of these hybrid models is critical in order for sociological rational choice theory to become central in sociology. The recent development of structural econometric models that are theory-driven in formulation indicates that we will fall behind economics unless we also push to develop empirically testable rational choice models.

The second category includes models that (re-)formulate theoretically-oriented stochastic process models as statistical models in order to estimate their parameters with empirical data. They include, for example,

1. Coale-McNeil model of marriage (1972), and

The Coale-McNeil model is based on certain general theoretical assumptions for the process of entry into marriage. Yamaguchi’s models reformulate some mathematical diffusion process models as accelerated-failure time regression models, and include a regression extension for a model developed by Herns (1972) for first marriage timing and for various models introduced by Diekmann (1989) for the analysis of diffusion processes. In the past, stochastic modeling of social and individual processes was a central component of mathematical sociology (e.g. Bartholomew 1982), but has now lost momentum mainly because it has been replaced by tools for the statistical analysis of panel data or event-history data. However, while it is certainly convenient to use standard statistical models, such models may not provide the rich information necessary in building sociological theories, because what we can do best with those models is to establish the absence or presence of a causal effect where we can adequately control for
selection bias into time-varying covariate states, without explaining the mechanism of causality. In order to develop a theory on social mechanisms (Hedstrom and Swedberg 1998), we need to combine evidences of causality with formal modeling of stochastic processes to strengthen the linkage between mathematical sociology and empirical research, not only to predict an occurrence of the event or a pattern of changes in states with some regularity but also to explain why such regularities exist. By relying on standard statistical models that are conveniently available from statistical software packages, we are compromising our theoretical concerns by not incorporating consideration for the mechanism into our formal analytical models.

The reason why the use of stochastic process models is being replaced by statistical methods for the analysis of panel or event-history data is not only due to convenience: statistical models permit the use of covariates and provide useful information. Hence, in order to compete against statistical models, mathematically-derived stochastic models must have a parameter for heterogeneity as the key element of the model so that covariate effects on the parameter can be estimated — that is, they must be extended into hybrid models. Compared with an extension of a rational choice model into a model that can be tested with survey data, this extension of a stochastic process model is a relatively easy task, given that a theoretically derived stochastic model is already formulated. What we need to do in order for the hybrid model to be constructed and applied to survey data is to identify a parameter in the stochastic process model for which the analyst derives benefits by assuming its heterogeneity in the population, and express the parameter as a parametric function of observable covariates. The method of parameter estimation may remain an obstacle depending on the complexity of the model, however.

3. Conclusion

There are multiple ways to link mathematical sociology and empirical research, some traditional and others new. It is very critical to strengthen linkages between mathematical modeling and empirical research in order for mathematical sociology to become more influential in sociological theory than it is currently. Generally speaking, mathematical sociology has three major advantages over “non-mathematical sociology.” One is that it can eliminate, or at least significantly reduce, potential ambiguities of concepts and descriptions of social mechanism associated with verbal expressions. Secondly, it permits a formal approach for introducing a micro foundation, which is rational choice and/or its modification, into the construction of theory for macro-sociological phenomena, which is what Coleman was concerned in his later years. Thirdly, it can model stochastic processes to describe social and individual processes,
which is what Coleman was concerned in his earlier years. The major obstacle to making mathematical models realistic has been the necessity of introducing institutional and/or structural effects on choices and processes. However, it seems that statistical techniques such as statistical modeling of social networks and statistical modeling of social interdependence are becoming ripe for incorporation of those structural considerations. There is now an increased potential to strengthen the linkage between mathematical modeling and statistical modeling via the construction of what I refer to as hybrid models, which will allow a test of our theoretical models with empirical data more directly than through the use of standard statistical analysis. Even when we retain a division of labor such that mathematical models are employed to derive empirically testable hypotheses and the data analyses are done separately to test them, we need more studies that represent what I refer to as the integrated division of labor, with which the use of mathematical models leads to new empirical findings predicted from the theoretical model, and thereby strengthens the empirical ground of the theory, rather than provides an alternative explanation for the social mechanism of facts that are already well known.

References


理論と方法

(Received July 1, 2005/ Accepted September 13, 2005)