19

JOEL E. COHEN
HARVARD UNIVERSITY

Aping Monkeys with Mathematics

This chapter offers evidence that as a means of understanding the functional and evolutionary biology of the primates, mathematical models are at least worth their weight in paper. The basic idea of mathematical models and the possibilities and the limits of their usefulness are outlined in Section 1, "Mathematical Models." As morphological arguments for evolutionary conclusions have become better defined, the use of mathematical models in morphology has increased. This trend is not without difficulties (as examined in Section 2, "Morphology in Evolutionary Arguments"). As behavioral and social characteristics of animals have appeared in evolutionary arguments, mathematical models are needed even more to help keep ideas dynamic (rather than typological), clear, and consistent (as explained in Section 3, "Behavioral Taxonomy"). Recent studies of casual social groups among human and nonhuman primates exemplify an explicit mathematical characterization of one aspect of social life. This characterization may provide a useful base for comparative evolutionary inferences (detailed in Section 4, "Monkeys en Masse").

The data used in quantitative studies of social groups can be related to frequencies of other aspects of behavior (asymmetric relations such as dominance, symmetric pairwise relations such as play partnership, and individual behaviors) by representing data in multidimensional contingency tables (in Section 5, "Data Structures and Models for Social Behavior"). The frequencies of behaviors obtained from multidimensional contingency tables, and the frequency distributions of choice of habitat, may be interpreted functionally on the basis of recent mathematically formulated discoveries in the psychology of operant conditioning (discussed in Section 6, "Frequencies of Behavior"). But most of the use of mathematical models in

studying the natural behavior, sociology, and ecology of primates (human and nonhuman) remains to be done (see Section 7, "Math and Aftermath").

Section 1. Mathematical Models

Like Moliere's gentleman who learned at an advanced age that he had been speaking prose all his life, those who claim ignorance of mathematical models may be surprised to discover that they have used them ever since elementary school (even without benefit of the "new math"). This recognition is important because insight into simple mathematical modeling can serve as a bridgehead to the understanding of much more complicated modeling.

By way of introduction to simple mathematical modeling (Cohen 1970), consider the two statements, "One plus one is two," and "One apple plus one apple makes two apples."

The second statement is an empirical statement about the world of experience. Once we agree on a procedure for counting, if we place one apple next to another apple we will both count two apples. The first statement is an arithmetical theorem which can be proved from simpler axioms along with many interesting results such as "Two times two is four."

Mathematical modeling is the act of imagination which proposes a connection between mathematical statements embedded in a mathematical structure, and some empirical observations. The difference between mathematics and mathematical modeling is like the difference in points of view one can have toward a microscope, which can be intrinsically interesting, or which can be a tool with which one extends one's perceived world.

In full flower, a mathematical model exemplified by "the laws of elementary arithmetic apply to apples and the way we count them" leads to such nonobvious predictions as "32,479 apples plus 90,503 apples makes 122,982 apples." Although this prediction has probably never been checked empirically, the behavior of small numbers of apples conforms so well to the elementary assumptions of arithmetic that few would doubt it. But the conscientious mathematical modeler cannot rest with proposing, generalizing, and verifying his model in special cases. He must also search for the limits of its validity. After observing apples in large quantities (or making calculations about the strength of materials), the modeler should point out that when more than one hundred thousand apples are gathered together, some of them become applesauce. Ordinary arithmetic does not then apply.

Nor does the truth of a model (arithmetic or any other) guarantee that the model is of interest. If instead of apples we put twenty monkeys together, the arithmetical details may be of less interest than the formation of a troop. And if we put two troops together, the number of troops (though

arithmetically predictable) may be of less interest than the interactions between them.

Whether a particular mathematical model interests a reader depends in part on the temperament of the reader, in part on the skill and insight of the modeler, and in part on the cooperation of nature.

Mathematical models are important in the study of primate evolution when they assist scientists to clarify what they believe about the phenomena they study. By thus sharpening the beliefs and expectations of scientists, mathematical models help to confirm or reject their assumptions and hypotheses. Mathematical models may facilitate finding explanations because they permit "what-if" experiments. If certain simplifying notions that are embodied in a mathematical model lead to qualitatively wrong predictions, then the scientist can abandon those bad notions. And even if they never lead to full explanations, mathematical models can clarify how much of the phenomena has been comprehended and expressed, and how good the resulting approximations are.

But mathematical models are not always necessary. Where ordinary concepts and everyday chains of logic suffice to give intellectual comfort with familiar kinds of materials, there is no scientific virtue in mathematical formalism. Three kinds of events can upset this amiable state of affairs: (1) a rise in the standard of intellectual comfort (as exemplified by other chapters in this book), so that what was once acceptable as an explanation is no longer adequate; (2) an extension of study to new kinds of materials where ordinary concepts and everyday logic are no longer competent (as in the study of "social evolution"); and (3) initially unrelated conceptual inventions, often new mathematics, which change the setting and illumination of familiar materials (see sections 5 and 7 to follow).

Section 2. Morphology in Evolutionary Arguments

The very progress of the study of primate evolution has created a need for mathematical models.

A fair caricature of classical morphological arguments in primate evolution may be provided by arranging four mandibles, one each from *Parapithecus, Propliopithecus, Pliopithecus,* and a modern gibbon, in a row (Clark 1957, p. 91). By looking at this arrangement, the reader is supposed to conclude that the modern gibbon mandible evolved from the smaller but similarly shaped *Pliopithecus* mandible, which in turn evolved from the still smaller *Propliopithecus* mandible, which originated from the smallest mandible, that of *Parapithecus.* But the graphic presentation of information does not make explicit at least two important steps which precede this conclusion.

First, the reader and author of the figure must assume that mandibles are of evolutionary significance and that, unlike certain other morphological

complexes in the organisms, the mandible is reliably and importantly related to the life, form, and evolution of the whole animal. Then both reader and author must characterize each mandible in a way that permits comparison among them, perhaps according to size, particular shapes and relative proportions of mandibular components, and specifications of the distribution of bone density. This characterization of the mandibles in a way that makes them comparable is actually the process of model construction; it is the abstraction from the complexity of real objects to an intellectually tractable description of what matters.

Second, the reader and author construct transformations of their characterizations of each mandible into those of the next mandible. The reader's sequence of transformations must be internally consistent (so that the step from *Parapithecus* to *Propliopithecus* is not too different by some measure from the step from *Pliopithecus* to *Hylobates*). Further, these transformations must be consistent with additional knowledge that the author and reader possess about the animals and evolutionary principles. Boorman (1970) offers an enlightening discussion on measures of distances between complicated objects.

Because, in classical morphological arguments, so much of the process from evidence to conclusion is private, there is enormous room for the author and his readers to arrive at different conclusions. If I think that only the shape and size of the canines significantly affect diet and evolution, while you attach much more importance to molars and mandibular size, and a third person believes that the scapula really matters and the mandible is not worth bothering with, then each of us will characterize the evidence differently. If I have one belief about phylogenetic allometry and you have another, even starting from the same characterizations we may differ in which evolutionary transformations we employ as standards of distance.

In spite of the great room for disagreements in the abstract, in fact substantial agreement has obtained in classical morphology. The whole mandible is one important determinant and consequence of diet and it is a morphological concomitant of other evolutionary changes. Progressive increases in its size without drastic changes in its shape constitute a plausible evolutionary sequence.

A great part of the rise in the intellectual standard of living among primate morphologists has been devoted to making explicit and otherwise improving procedures for characterizing (modeling) the objects of their studies. D'Arcy Wentworth Thompson superimposed square grids (Cartesian coordinates) on drawings of forms, and showed how they could be simply distorted to produce drawings of apparently related forms.

Oxnard (1969c, pp. 75–76) observed:

The shape of a biological specimen may be represented by a series of measurements of different kinds taken on the specimen. One way, therefore, of repre-

senting the specimen is as a single point located in a multidimensional space. The coordinates of this point are the actual values of each of the measurements, and the many dimensions are the many different measurements. . . . [See also Howells, Chapter 5.] But each original measurement does not necessarily give completely new information about a particular specimen; for instance, a second measurement of the radius of a circle tells us nothing new at all about the circle.

A mathematical technique called canonical analysis provides a way of combining measurements into a new set of dimensions such that each additional dimension in this new set provides the maximum increase in information (by some measure of information) about the shape of the specimen. Canonical analysis may show that a much smaller number of dimensions suffices to characterize specimens, though these dimensions would not have occurred to a naive observer. Given improved characterizations, improved speculations about how evolution transforms one to another are possible (Oxnard 1969c, pp. 92–95).

Multivariate statistical methods have also been used to reveal morphological relations among populations of men. Howells (1970a) found discriminant functions which would characterize efficiently the differences among 17 populations of contemporary human skulls. He used these functions (1970b) to assist in the interpretation of skulls of Mount Carmel man.

These multivariate statistical methods improve earlier eyeball "guesstimations" of form. But these methods can cause problems which call for more refined and different mathematical modeling. Here is an example:

Suppose we are investigating a polygenic character. Suppose also that the number of genes (additive polygenes, not loci) for this character is proportional to the length of time that some spherical morphological feature grows. But the measured aspect of that feature is its volume. In this case, the number of genes for the feature will be related to the measurement of the feature as x is to x^3, and in general as x is to $f(x)$.

Suppose that for five known populations, numbered 1, 2, 3, 4, and 5, the numbers of genes for this character are, respectively and exactly, 1, 2, 3, 5, and 6, and hence that the phenotypic measurements (which are the only direct measurements we can make) are 1, 8, 27, 125, and 216. We observe now a new sixth population with phenotypic measurement 343. The linear phenotypic distance between population 6 and population 5 (namely, $343 - 216 = 127$) is greater than the linear phenotypic distance $125 - 1 = 124$ between population 4 and population 1. We therefore conclude from linear analysis of phenotypic measurements that populations 1 and 4 are more closely related than are populations 5 and 6.

But if our concept of relatedness is to be based on genetic similarity, it is clear that two populations (5 and 6) which differ in only one polygene are more closely related than populations (1 and 4) which differ by three polygenes. Further, on the genotypic level, populations 1, 2, and 3 are

about as closely related as are populations 4, 5, and 6; but on the phenotypic level, the discovery of population 6 appears to make populations 1, 2, and 3 more closely related, relative to the whole scale of variation.

If, as Norbert Wiener purportedly said, what we want are measurements that will stay put when our backs are turned, the accessible phenotypic measurements are clearly less satisfactory than the inaccessible genotypic ones. One role for theory is to transform measurements, e.g. $f(x)$, into evidence, x, which will not require a new view of the world to accommodate every new experience. The mechanized algorithms of linear statistical analysis may be very useful aids in preliminary probes designed to discover what theory must accomplish, but they are not adequate theories.

Section 3. Behavioral Taxonomy

In the study of behavior, the temptation to affirm that seeing is believing is no weaker than in the study of morphology, and the need for theory to relate observations to some underlying invariants is no less urgent.

In 1898, Charles Otis Whitman emphasized behavioral approaches to evolution: "Instincts and organs are to be studied from the common viewpoint of phyletic descent." His successor Konrad Lorenz (1958) put it thus:

> As phylogenists, Whitman and Heinroth both sought to develop in detail the relationship between families and species of birds. To define a given group they had to find its "homologous" traits; the resemblances between species which bespeak a common origin. . . . Behavior, as well as body form and structure, displays homologous traits.

Marrying behavior and morphology in the study of primate evolution, Tuttle (1969b) studied the hands of the great apes. Although using no mathematical models, he characterized the role of the hands in locomotion by the bone structures in modern and fossil apes, and by the behaviors of modern apes in knuckle-walking, fist-walking, and modified palmigrade walking. He considered plausible transformations of these descriptions and showed them to render unlikely some suggested evolutionary sequences leading to man. Remarkably, Tuttle (1969b, p. 957) declined to claim that knuckle-walking devolved from fist-walking and not modified palmigrade locomotion, or vice versa. Even without a mathematical model, Tuttle obtained what is often one of the chief benefits of a mathematical model: a demonstration that present data are insufficient to choose among different possible transformations leading to a given result.

Detailed studies of one aspect of primate morphology or behavior, such as those of Oxnard and Tuttle, have the virtue of specifying the procedure for measuring the dimensions of the object of study with sufficient public detail that another worker can check the measurements on his own material or adapt the procedure to new kinds of materials.

Studies in similar detail of the behavior of human individuals in a social setting led Bales (1955) to propose four broad categories of talk: positive emotional reactions, attempts to solve a problem, questions, and negative emotional reactions. In observing primates whose language, if it exists, is not understood or difficult to observe, only the first and last categories of acts seem applicable. These categories reappear in the behavior of rhesus monkeys. Altmann (1968, pp. 62–63) first defined a detailed behavioral repertoire for rhesus monkeys, and then looked inductively for groupings in the sequences of those behavioral elements. He found these groups: primarily affinitive and mild tactile signals; primarily agonistic signals; primarily sexual signals of males; primarily sexual signals of females; play; suckling and weaning; aggressive chases; carrying of infants; and two miscellaneous agonistic patterns. Applied to other species, Altmann's detailed procedure might lead to comparable groupings of behavior patterns.

Section 4. Monkeys en Masse

Even some of the most systematic attempts to advance from individual behavior to social systems return to a level of discourse comparable to that of morphology before D'Arcy Thompson, because they follow the strategy of arranging homologous bits of (social) anatomy in suggestive sequences as Clark (1957, p. 91) arranged four mandibles to represent a hylobatid lineage.

Crook and Gartlan (1966) named five "adaptive grades" of primates, moving from forest dwelling through forest fringe and tree savannah to grassland or arid savannah, with concomitant changes in diet, diurnal activity, size of troops, reproductive units, male mobility, sexual dimorphism, and population dispersion. In this typology, Crook and Gartlan specified "size" of troops (which they called "groups") in successive grades as "usually solitary, very small groups, small to occasionally large parties, medium to large groups, . . . [and] large groups. . . ." Here any actual measurement of the size of troops is implicit. The introduction of a Cartesian grid for all visible dimensions or aspects of the primate adaptations is clearly not even contemplated.

Kummer (1968) and Altmann and Altmann (1970) provide explicit measurements of many different aspects of social patterning in baboons. These cross-sectional studies of single species run no risk of overlooking aspects of behavior or ecology which lack apparent homologues in other species. Perhaps in reaction to global and glib comparative essays less scrupulous than that of Crook and Gartlan (1966), these studies indulge sparingly in comparative or evolutionary speculation. The detail with which their observations are made and reported offers the possibility of making similar observations on other primate species, and then making potentially credible comparative speculations involving explicit mathematical models.

Rather than studying in detail the social behavior of one entire species, and leaving comparative work for later, another approach is to select one aspect of social behavior and study it comparatively in great detail, leaving other aspects of social behavior for later. Such a narrow but longitudinal approach may provide suggestions and indications, early in the course of study, about those aspects of social behavior which have evolutionary interest or which correlate in interesting ways with other characters of evolutionary interest. If both the longitudinal and cross-sectional approaches are well based in careful observation, and if the principal elements of each species' sociobiology have evolutionary meaning and therefore comparative homologues, the two approaches should lead to similar evolutionary inferences.

One narrow aspect of primate social behavior which has been studied comparatively by means of mathematical models is the formation of casual or spontaneous social groups (Cohen 1971; and in press).

In the early 1950s, John James observed the sizes of freely forming small groups of human beings in a variety of situations. James defined freely forming groups as "those whose members are relatively free to maintain or break off contact with one another; that is, they are ones where informal controls on behavior are at work and spontaneity is at a maximum." He included in his observations only those "groups in which the members were in face-to-face interaction as evidenced by the criteria of gesticulation, laughter, smiles, talk, play, or work. Individuals who merely occupied contiguous space were not counted as members of a group." Goffman's (1963) far subtler analysis of social interactions in public places makes it clear that James' definitions and approach overlook a great deal, but does not render them less useful for present purposes.

The full frequency distributions of the nearly 18,000 groups James observed in 21 different situations were published in Coleman (1964b, pp. 368–373). Situations observed included pedestrians in Eugene, Oregon on a spring morning; shopping groups in two Portland, Oregon department stores; and play groups in Eugene in the spring on the playgrounds of 14 elementary schools (directed or organized play was not included in the observations).

In 1961 Coleman and James observed that a Poisson distribution with the zero value truncated described nearly all of the observed frequency distributions very well. Coleman proposed a set of assumptions about a group's probabilities of transition from one size to another which led to a prediction at equilibrium of a truncated Poisson distribution of group size.

The next year Harrison White pointed out that Coleman's model treated each group as if its behavior were independent of the sizes of all the other groups in the system of groups (in the playground or in the department store). White proposed several "sociological" models in which the behavior of particular groups depended on the number of groups of other sizes in

the system. These models differed in assumptions about the flow of individuals between the system and its environment, the flow of members among the groups, and the dependence of rates of arrival to and departure from groups on sizes and numbers of groups. In spite of substantial differences in assumptions, White claimed to show that all the models led to an equilibrium distribution of group size given by the truncated Poisson. As did Tuttle in considering alternate pathways for the evolution of knuckle-walking, White concluded that "more elaborate empirical investigations . . . will be needed to test the applicability of and to discriminate among simple stochastic models for the circulation of members among casual groups."

In 1964 Goodman criticized the mathematics of Coleman and White, but concurred in the conclusion that a variety of models could lead to the same equilibrium distribution. He proposed yet another model.

Thus James discovered a striking regularity in the social aggregations of large collections of people, although only 18 of James's 21 observed frequency distributions were well described by the truncated Poisson distribution. Coleman, White, and Goodman provided several different explanations, each accompanied with a plea for further empirical research.

These stochastic models for systems of social groups appeared at the same time as the first of the contemporary spate of primate field studies. The coincidence raised the hope of finding primate field data on casual social groups which, first, would show whether the Poisson regularity held for primate species other than man; and second, would discriminate among the existing models leading to the Poisson distribution.

In 1967, Stuart Altmann directed me to the last table in Thomas T. Struhsaker's doctoral dissertation on vervet monkeys in East Africa. For each night of observation in this table, Struhsaker recorded which individual monkeys of a particular troop were sleeping together in one tree top, which monkeys if any in another, which if any in a third, and so on.

These sleeping groups are not casual groups within James' definition because the vervets cannot leave in the middle of the night, and during sleep there may be hardly any interactions. Nevertheless I assumed that the groups represented a snapshot of the state of the system of groups within the troop at the time that the monkeys ascended into the trees for the night. The frequency distribution of group size was clearly not Poisson. But it approximated a truncated negative binomial.

Now, either the process of forming human casual social groups and the process of forming vervet sleeping groups are different, or both are manifestations of some common underlying process (at least as far as their size is concerned). I preferred the latter alternative, and invented a family of stochastic models, called linear one-step transition (LOST) models, which depend on two ratios, a/d and b/d, of three parameters, a, b, and d. Parameter a is a proportionality constant which measures the rate at which individuals join groups, independently of the size or identity of the individ-

uals in the groups. Parameter b is a proportionality constant which measures the rate at which individuals join groups in proportion to the numbers of individuals in those groups (but independently of the identities of the individuals in the groups). Parameter d is a proportionality constant which measures the rate at which individuals already in groups make independent decisions to leave their groups and become isolates. For positive values of b, these models predict a truncated negative binomial distribution of group size; and for $b = 0$ they predict a truncated Poisson distribution of group size.

Thus, these models encompass Struhsaker's data on vervet sleeping groups and those data of James previously described by the truncated Poisson distribution. Reexamination of James' three distributions which did not fit the Poisson showed that two of them also could be fitted by the truncated negative binomial.

These data were not sufficient to test whether the detailed assumptions of the LOST models concerning arrivals and departures were true, but the detail in Struhsaker's observations was sufficient to rule out a variety of other models that had been proposed as mechanisms leading to the negative binomial distribution.

To test directly the detailed mechanisms of the LOST models, I observed the free play of four-year-old humans in the Cambridge Nursery School. There appeared to be no large differences between what the LOST models assume about the dynamics of group formation, and how the actual sizes of play groups changed over 30-second intervals. This surprising result did not show that the same dynamics were at work in any species of non-human primate; it merely failed to rule out the possibility that the LOST dynamics were at work among humans.

I then wanted to see whether monkeys' undirected socializing could be described by the same models as human socializing. In 1969 Stuart Altmann provided me with an opportunity to answer my questions by joining his field project on baboons in Kenya. The setting of this study, as of 1964, is described by Altmann and Altmann (1970). Details on my observations appear in Cohen (in press). In brief, of the seven troops I observed for sufficient periods, only one had a group size distribution which fit the predicted distribution poorly (using less than fully efficient methods of estimation and a test of goodness of fit with artificially increased power).

The generally acceptable agreement between the observed frequency distributions of group sizes in men and monkeys and the fitted distributions shows that a single family of models may suffice for present purposes. The dynamic assumptions of the models have been tested only on nursery school observations. Since the models characterize those situations they describe by the two numbers a/d and b/d, the net result of the studies to date can be summarized by a list of the primates observed and their associated values of a/d and b/d.

*Table 19–1. Median values of the parameters of the linear one-step
transition (LOST) models for various primates*

	a/d	b/d
Vervets (Struhsaker)	1.15	0.66
Yellow baboons	0.12	0.16
Human four-year-olds	0.33	0.10
Mixed humans (James)	0.86	0

The median values of the parameter ratios for baboons in Amboseli were
chosen as typical. The vervet parameter values are based on Struhsaker's
observations during nights when all members of the troop were noted. The
nursery school figures are an average of the estimates obtained on different
days of observation weighted by the number of observations on each day.
Finally, the figures based on James's observations are the medians of Cole-
man's estimates for the 18 cases where the truncated Poisson distribution
was descriptive and my estimates for the two other cases where the negative
binomial distribution was descriptive.

The systems of casual social groups characterized by the LOST models
bear the same relation to the whole social behavior of a primate species that
the hand or perhaps a single digit do to the total morphology of an orga-
nism: they are important, and not the whole picture. The models show
one way to characterize, perhaps usefully, a complicated kind of social
behavior by a small number of parameters. Characterizing a primate social
system by parameters embedded in a predictive model is self-validating:
one need take the characterization and the parameter values seriously only
when the predictions of the model are confirmed. If the equilibrium dis-
tribution of group sizes in a system of casual social groups is neither a
truncated negative binomial nor a truncated Poisson, then the parameter
values—no matter how operationally defined—are no longer providing
useful information.

The characterization of social systems by operationally defined parameters
embedded in models gives concrete meaning to the phrase, "the evolution
of social systems." Such evolution is simply the trajectory in time of the
parameter values typical of a taxon and situation, within the span of evolu-
tionary time that the models are useful.

The estimates of b/d in the rightmost column of Table 19–1 decline as
one reads down the table. Though the difficulties of evolutionary inference
from the behavior of presently existing species are manifold, I wonder if
there is any evolutionary meaning to the decrease in the role of individual
attraction in relation to individual departures (b/d) and the ascendancy
of group attraction in relation to individual departures (a/d). A theory of
this phenomenon, if it exists, is evolutionary theory, and can be distinguished
from the theory of the social system itself.

Section 5. Data Structures and Models for Social Behavior

The studies of casual social groups just described report observations on the frequency distribution of the size of groups engaged in a particular class of activities (informal socializing).

Transaction flows are another common form of reporting quantitative observations of social behavior. In transaction flows, an actor and an object are designated; for example, a speaker and a listener (Bales 1968), a dominating chicken and a dominated chicken (Guhl 1956), or a threatener and a threatened subject. Transaction flows are usually reported as a matrix in which the rows correspond to actors and the columns correspond to objects. The number in the ith row and jth column reports the frequency with which individual i acted upon individual j. In such matrices, the diagonal is zero or undefined. The statistical analysis of such matrices, assuming certain probability models, has been treated by Bishop and Fienberg (1969), Fienberg (1969), Goodman (1968), and Wagner (1970). When the individuals are lumped into classes such that one member of a class may act upon another member of the same class, the diagonals need not be zero (e.g. Altmann 1968).

A third form of recording quantitative observations of social behavior consists of matrices in which rows and columns are identified as before but in which the individual entries record the frequency of some conjoint action, such as the frequency with which monkey i and monkey j slept together in the tree tops (Struhsaker 1967b, p. 112). Clearly this relation is symmetric: the entry in row i and column j equals the entry in row j and column i. When columns and rows are identified with individuals, the diagonal may be left undefined or may record the frequency of the action (if meaningful) by the isolated individual. The half-matrix above the diagonal suffices to record the pairwise frequencies when rows and columns are identified with individuals. When individuals are pooled into classes, the diagonal is required as well.

A fourth form of recording quantitative observations of social behavior (to which I will return in Section 6) constitutes a frequency distribution on patterns or elements of behavior. This frequency distribution tells, for example, how often all individuals chased, fled, slept, or played. Altmann (1965) gives a frequency distribution on a behavioral repertoire of 120 elements.

A fifth method that could be employed to record quantitative observations of social behavior is the multidimensional contingency table. Each dimension is associated with an individual in the social system. Each category along the dimensions is associated with an element in a behavioral repertoire. This repertoire must be defined so that its categories are mutually exclusive. This requirement excludes a repertoire which permits an animal, for example, to be groomed and to be grooming at the same time if these

are considered two different elements of the repertoire. Multidimensional contingency tables have recently been used to analyze lizard feeding behavior and interspecific interactions in ecology (Schoener 1970). A simple guide to their statistical analysis in this context is available (Fienberg 1970). To my knowledge, these tables have not been used to record social behavior.

The information contained in the first four forms for recording observations of behavior can be derived by arithmetic from this fifth form. Further, a model for the probabilities of observations in the cells of the multidimensional contingency table implies models for each of the other four arrays of observations, by the same arithmetic operations performed on the predicted probabilities. In this sense, a model for the multidimensional contingency table is an underlying model for the other kinds of observations. In particular, a model of independence in conjoint action as recorded in symmetric matrices (the third form above) and the Poisson or negative binomial theoretical distributions of group size (as predicted by the LOST models of the previous section) can both be derived from a single, simple partial model for the multidimensional contingency table. (The much more difficult problem of the effects of collapsing multidimensional contingency tables when a model of independent dimensions is not assumed is being studied by Bishop [1971].)

To keep the example simple, suppose we have a primate troop of only three individuals that do only four things: chasing, fleeing, sleeping in the troop's roost, and playing in the troop's playpen. All individuals in the troop are presumed to be in view with each observation of the troop. (This assumption avoids all the difficult problems of sampling.) At any time, each of the animals is engaged in precisely one of these activities. An animal can be engaged in chasing if and only if there are one or more animals fleeing from it, and an animal can be fleeing if and only if there are one or more animals chasing it, and no animal can be both chasing and fleeing at the same time. (In the common sense of the terms "chasing" and "fleeing," an animal may be both chasing and fleeing at the same time; but we require the categories to be defined as disjoint.) Any binary social relation, such as talking and listening, which satisfies the same properties may be substituted for chasing and fleeing. Arbitrary n-ary relations may be recorded in the same contingency table format so long as there are at least n individuals in the troop.

Then observations of the entire troop's social activities may be entered in the three-dimensional array with four categories in each dimension, as shown in Table 19–2. Here the letters $a, b, c, \ldots x, y, z, A$ are numbers of observations. Roman numerals I, II, III identify the three individuals, and (1), (2), (3), (4) identify chasing, fleeing, sleeping, and playing, respectively.

The zeros are logical zeros. For example, the entry in the third row and third column of the first panel is zero because if animal I is chasing (1), and animal II is sleeping (3), by definition of "chasing" animal III must

Table 19–2. A multidimensional contingency table to record social behavior in a troop of three individuals (I, II, III), each having four patterns of behavior [(1), (2), (3), (4)], entries a, b, c, . . . y, z, A being observed frequencies

		II(1)	(2)	(3)	(4)
	III(1)	0	a	0	0
I(1)	(2)	b	c	d	e
	(3)	0	f	0	0
	(4)	0	g	0	0
		II(1)	(2)	(3)	(4)
	III(1)	h	i	j	k
I(2)	(2)	l	0	0	0
	(3)	m	0	0	0
	(4)	n	0	0	0
		II(1)	(2)	(3)	(4)
	III(1)	0	p	0	0
I(3)	(2)	q	0	0	0
	(3)	0	0	r	s
	(4)	0	0	t	u
		II(1)	(2)	(3)	(4)
	III(1)	0	v	0	0
I(4)	(2)	w	0	0	0
	(3)	0	0	x	y
	(4)	0	0	z	A

be fleeing (with frequency d, row (2)), and cannot also be sleeping. All the other zeros are argued similarly.

Table 19–3 illustrates the four derived data formats, in the order in which they were first described. The entries in these arrays show how they were derived from the counts in the multidimensional contingency table (Table 19–2). Part I of Table 19–3 gives the frequency distribution of sleeping group sizes and the frequency distribution of play group sizes. For example, the frequency with which exactly two animals were observed playing together is $u + y + z$. The frequency with which all three animals slept in one group is r.

The frequencies with which particular pairs of animals engaged in chasing and flight are given in part 2 of Table 19–3. Thus animal I chased animal III with frequency $b + c + d + e$, whether or not animal I also chased animal II or was joined in chasing III by animal II.

The frequencies with which each pair of animals slept together and played together are given in part 3 of Table 19–3 regardless of the presence or absence of the remaining animal in the troop. Thus animal II and animal III slept together with frequency $r + x$.

Finally, the frequencies with which the acts of chasing, fleeing, sleeping, and playing were observed are given in part 4 of Table 19–3. Since A is the frequency with which all three animals were observed playing at the same

Table 19–3. Four frequency distributions derived from the multidimensional contingency table (Table 19–2): 3.1, group sizes; 3.2, transaction flows; 3.3, pairwise conjoint actions; 3.4, behavior patterns

3.1

Size of Groups	Frequency of Sleeping (3) Groups	Frequency of Play (4) Groups
0	$a+b+c+e+g+h+i+k$ $+1+n+p+q+v+w+A$	$a+b+c+d+f+h+i+j$ $+1+m+p+q+r+v+w$
1	$d+f+j+m+u+y+z$	$e+g+k+n+s+t+x$
2	$s+t+x$	$u+y+z$
3	r	A

3.2

		Fleeing (2)		
		I	II	III
	I	0	$a+c+f+g$	$b+c+d+e$
Chasing(1)	II	$h+1+m+n$	0	$b+1+q+w$
	III	$h+i+j+k$	$a+i+p+v$	0

3.3

Sleeping (3) pairs	I	II	III	Playing (4) pairs	I	II	III
I	0	$r+t$	$r+s$	I	0	$y+A$	$z+A$
II	$r+t$	0	$r+x$	II	$y+A$	0	$u+A$
III	$r+s$	$r+x$	0	III	$z+A$	$u+A$	0

3.4

Behavior	Frequency
(1)	$2a+2b+c+d+e+f+g+2h+i+j+k+1+m$ $+n+p+q+v+w$
(2)	$a+b+2c+d+e+f+g+h+2i+j+k+21$ $+m+n+p+q+v+w$
(3)	$d+f+j+m+y+z+u+2(s+t+x)+3r$
(4)	$e+g+k+n+s+t+x+2(u+y+z)+3A$

time, each such observation contributes 3 to the number of times which the act of playing was observed; similarly for the other entries multiplied by numbers.

The point to be made is that if the contingency table in Table 19–2 is available, all the other forms in Table 19–3 of describing social behaviors follow by arithmetic.

A model of quasi-independence which is widely used for the analysis of transaction flows (Bishop and Fienberg 1969, Fienberg 1969, Goodman 1968, Wagner 1970) attributes to each individual $i = $ I, II, III a chasing-strength $\alpha_{i(1)}$ and a fleeing-strength $\alpha_{i(2)}$, both of which are dimensionless probabilities between 0 and 1. The model predicts that the frequency with which individual i will be observed chasing individual j will equal $N\alpha_{i(1)}\alpha_{j(2)}$,

where N is the number of observations of the whole troop. For symmetric relations such as sleeping or playing, the matrices of expectations $N\alpha_{i(3)}\alpha_{j(3)}$ $= N\alpha_{j(3)}\alpha_{i(3)}$ and $N\alpha_{i(4)}\alpha_{j(4)} = N\alpha_{j(4)}\alpha_{i(4)}$ are symmetric. That there is no statistical or logical necessity for observations of pairwise interactions to be described by a model of quasi-independence is most convincingly demonstrated by the failure of real data to be so described (Altmann 1968).

To relate the symmetric model for a conjoint action, e.g. playing (4), to a Poisson or negative binomial distribution for the size of play groups, we construct a model for the $2 \times 2 \times 2$ contingency table in part 1 of Table 19–4, which is obtained from Table 19–2 by collapsing the four elements of the behavioral repertoire into two elements, namely, "playing" and "not playing." (When there are n animals in the troop, this censorship yields a 2^n contingency table for the frequencies of conjoint playing.) The model of independence, which gives each animal i a probability $\alpha_{i(4)}$, $i =$ I, II, III, of joining in play, independently of all other animals, assigns to each combination of animals the probabilities in part 2 of Table 19–4. When these are transformed to a symmetric matrix part 3 of Table 19–4 analogous to part 3 of Table 19–3, it appears that the probability that any animal is paired with any other animal in play equals the product of their respective probabilities of playing; hence that the model of quasi-independence is satisfied.

When the probabilities in part 2 of Table 19–4 are transformed to a frequency distribution of group size as in part 4 of Table 19–4, it appears that if all the animals' probabilities of play $\alpha_{i(4)}$ were equal to $\alpha_{(4)}$, then group sizes would be binomially distributed, with probability parameter $\alpha_{(4)}$ and with $n = 3$. It is well known that if n becomes large and $\alpha_{(4)}$ remains small so that $n\alpha_{(4)}$ is moderate, the binomial distribution approximates the Poisson distribution with parameter $\lambda = n\alpha_{(4)}$. If the animals in a large troop were divided into classes, the members of which played only with other members of the same class, and the probabilities $\alpha_{(4)}$ were distributed over the classes in an approximately gamma distribution, then the group size distribution for the whole troop would be approximately negative binomial.

This derivation of the Poisson and negative binomial distributions from the underlying model of independence presented in part 2 of Table 19–4 rests on the assumption that all $\alpha_{i(4)}$ are equal. This assumption may be relaxed. If the probabilities $\alpha_{i(4)}$ are distributed over the animals i in a beta distribution, then it may be proved that the distribution of group sizes will be hyperbinomial. As n gets large, for small group sizes and subject to certain reasonable constraints on the beta distribution (small variance), the hyperbinomial distribution approximates a binomial distribution. The details of this argument and the explicit form of the restrictions on parameters which make this limiting process valid may be found in Pratt, Raiffa, and Schlaifer (1965, ch. 9). This binomial distribution of group sizes yields

Table 19–4. Modeling the multidimensional contingency table in order to relate models of pairwise conjoint actions and group size distributions:* 4.1, Table 19.2 censored to show only playing (4) or not playing (−4); 4.2, model of independence for 4.1; 4.3, model for pairwise conjoint actions implied by 4.2; 4.4, model for group sizes implied by 4.2

4.1

		II(4)	II(−4)		II(4)	II(−4)
I(4)	III(4)	A	z	I(−4)	u	$g+n+t$
	III(−4)	y	$v+w+x$		$e+k+s$	$a+b+c+d+f$ $+h+i+j+l$ $+m+p$ $+q+r$

4.2

		II(4)	II(−4)
I(4)	III(4)	$\alpha_1\alpha_2\alpha_3$	$\alpha_1(1-\alpha_2)\alpha_3$
	III(−4)	$\alpha_1\alpha_2(1-\alpha_3)$	$\alpha_1(1-\alpha_2)(1-\alpha_3)$
I(−4)	III(4)	$(1-\alpha_1)\alpha_2\alpha_3$	$(1-\alpha_1)(1-\alpha_2)\alpha_3$
	III(−4)	$(1-\alpha_1)\alpha_2(1-\alpha_3)$	$(1-\alpha_1)(1-\alpha_2)(1-\alpha_3)$

4.3

	I	II	III
I	0	$\alpha_1\alpha_2$	$\alpha_1\alpha_3$
II	$\alpha_1\alpha_2$	0	$\alpha_2\alpha_3$
III	$\alpha_1\alpha_3$	$\alpha_2\alpha_3$	0

4.4

Size of groups	Probability that a play (4) group will have the size given
0	$(1-\alpha_1)(1-\alpha_2)(1-\alpha_3)$
1	$\alpha_1(1-\alpha_2)(1-\alpha_3) + (1-\alpha_1)\alpha_2(1-\alpha_3) + (1-\alpha_1)(1-\alpha_2)\alpha_3$
2	$(1-\alpha_1)\alpha_1\alpha_2 + \alpha_1(1-\alpha_2)\alpha_3 + \alpha_1\alpha_2(1-\alpha_3)$
3	$\alpha_1\alpha_2\alpha_3$

*In this table α_i means $\alpha_{i(4)}$, i's probability of playing.

the Poisson and the negative binomial distributions by the same processes as before.

Hence it is possible to go from a single underlying model as represented in part 2 of Table 19–4 to both the model of quasi-independence in pairwise interactions and the same equilibrium distributions of group sizes (Poisson and negative binomial) which are predicted from other (LOST) models. The multidimensional contingency table (Table 19–2) offers a way to unify apparently different quantitative approaches to the analysis of social behavior, and raises an obvious further question (as Stephen E. Fienberg has pointed out privately): What model of the full Table 19–2 will predict quasi-independence for part 2 of Table 19–3 and symmetric quasi-independence for symmetric pairwise activities such as part 3 of Table 19–3? The answer is instantly obvious to neither him nor me.

Section 6. Frequencies of Behavior

Why are the frequencies of behavior what they are? If the frequencies are presented as a summed distribution over the repertoire (e.g. part 4 of Table 19–3), then an initial procedure is to explain the frequencies by deriving them from some probability model of the underlying contingency table (Table 19–2). Thus if one is interested only in the frequency of playing, one could relate the frequency to the underlying probabilities $\alpha_{i(4)}$, $i = $ I, II, III of joining in play. This is only temporarily satisfying, and merely changes the question to: why are the probabilities $\alpha_{i(l)}$, $(l) = (1)$, (2), (3), (4) what they are?

One answer to this question may be based on physiological mechanisms in the animal which are triggered by its environment with a frequency corresponding to the frequency of the behavior. When a clear-cut trigger exists, as in many mating rituals, this answer can be illuminating. It is less satisfactory when animals behave in similar ways in a variety of circumstances and when, under similar circumstances, different individuals and species behave variously.

A response of evolutionary biologists is that when individuals of a species achieve behavioral equilibrium in an environment, the frequency distribution of observed behavioral patterns is that which is most adaptive, in the sense of maximizing Fisher's Malthusian parameter. For example, Altmann and Altmann (1970, pp. 198–201) attempt to explain a baboon troop's allocation of time to different quadrats of a habitat as one which maximizes the sum of the differences between benefits and costs (measured in terms of reproductive success) associated with each quadrat. Testing this approach requires a direct measurement of the benefits and costs associated with each quadrat occupied or behavior pattern observed. Field studies of primates are only beginning to provide such measurements.

However, direct quantitative evidence is available that pigeons, rats, and people in psychological laboratories match their behavior to the rewards for different acts provided by the environment. The following description is based on Herrnstein (1970; and in press).

Suppose a pigeon in a Skinner box has a choice of two disks to peck at. Suppose that for each disk there is a minimum interval of time (possibly different for each disk) such that as soon as the pigeon pecks after that interval it is reinforced with food, but if it pecks before the end of the interval it gains nothing; and suppose this interval fluctuates around some mean. This schedule of reinforcement is called a "variable interval" schedule and usually produces in pigeons a rate of pecking far greater than the rate of reinforcement.

Herrnstein showed that the pigeons matched their pecking to the rewards from each disk. For example, when the average rate of reinforcement from the left disk was 30 reinforcements per hour and the average

rate from the right disk was 10 per hour, then each pigeon delivered three-quarters = 30/(30 + 10) of its pecks to the left disk and one-quarter to the right. An extensive series of experiments led to the conclusion (Herrnstein in press) that "responding and reinforcement are related by a constant of proportionality, *k*. Reinforcement, however, is not measured absolutely, but as a ratio between the reinforcement conditional upon the response and total reinforcement." If $P_{i(l)}$ is animal *i*'s rate of performing acts of category (*l*), and $R_{(l)}$ is the reinforcement to acts of type (*l*), then Herrnstein found:

$$P_{i(l)} = \frac{k_i R_{(l)}}{R_{(0)} + R_{(1)} + \ldots + R_{(m)}}, l = 1, \ldots m. \tag{1}$$

The number of different acts visibly reinforced is *m*. $R_{(0)}$ is a residual reinforcement for acts not listed in the repertoire, akin to the animal's proclivity for self-generated entertainment without external reinforcement. It follows that for each animal *i* the probability $\alpha_{i(l)}$ of acts of type (*l*), *l* = 1, 2, . . . *m*, among all *m* overtly reinforced acts in the repertoire, is:

$$\alpha_{i(l)} = \frac{P_{i(l)}}{P_{i(1)} + \ldots + P_{i(m)}} = \frac{R_{(l)}}{R_{(1)} + \ldots + R_{(m)}}.$$

Hence there is a matching between relative rates of response and relative reinforcement.

An important implication of expression (1) noted by Herrnstein is:

Contrary to intuition, responding may therefore be more or less indifferent to the reinforcement it produces. If the response's reinforcement is a large part of the total reinforcement, which is to say that the reinforcement ratio in equation [1] is close to 1.0, then responding will stay close to the value of *k*. And conversely, if the response's reinforcement is only a small fraction of the total reinforcement, then the response will be quite sensitive to variations in its reinforcement, in the limiting case being directly proportional to it.

The confirmation of the generalization summarized by equation (1) in more than 50 experiments suggests that it also might be a valuable aid in interpreting frequencies of behavior outside of Skinner boxes as attempts to match environmental reinforcements.

The baboons' occupancy of different quadrats is not a behavior with a rate, but a choice with duration. An experiment which shows that the same regularity, Equation (1), describes such choices is described by Herrnstein:

Pigeons were given the chance to choose between blue or amber illumination of the experimental chamber. One peck at a disk changed it to amber if it was blue, and vice versa. Every now and then, the pigeon (who was hungry) was given a bit of food irrespective of its responding. The rate of feedings depended on the color of illumination. At any moment, the pigeon could switch

from the prevailing color to the other. Except for the lack of a response requirement, the two schedules were variable intervals running concurrently, so that it was advantageous for the pigeons to switch back and forth to collect the feedings that were coming due. . . . the pigeons kept the proportion of time spent in a given color of illumination equal to the proportion of reinforcements obtained therein.

If the various quadrats of a baboon's habitat provide resources in an approximation to a variable-interval schedule, then the observed durations of occupancy of the different quadrats may indeed be matching the net gains from the quadrats, as the Altmanns have suggested. The finding of the experimental psychologists shows that such matching occurs in the laboratory and defines the precise form of the distribution of net benefits which should be looked for in the field. That the net gain from different quadrats, or from different patterns of behavior in the field, in fact matches the duration of occupancy or frequency of performance, remains unproved.

Section 7. Math and Aftermath

In addition to those already mentioned, other mathematical techniques are ripe for a fruitful union with empirical techniques and observations within the next few years. Moreover, the growth of mathematical models peripheral, but nonetheless related, to primatology may offer suggestive leads.

One area of primatology ready for a union of existing models and data is the study of dominance. Many data of varying degrees of detail (Kawai 1958; Itani et al. 1963; Mizuhara 1964; Bernstein and Draper 1964; Bernstein 1968b, 1969, 1970; Uyeno 1967; Delgado 1967; Vandenbergh 1967; Sade 1969a; and others) have never been compared systematically with theory (e.g. Landau 1968; Boorman 1970, and in preparation). Mathematical models recently developed in the theories of preference and measurement (Roberts 1969) seem readily translatable into sociological situations where transitivity is absent. But their relevance remains unexplored. Structural models of roles in human societies may well offer insight into primate social relations when simple dominance models fail (Lorrain and White 1971).

Second, the population genetics of primate troops is now becoming accessible to study through the conjoint development of improved methods of assessing and understanding the genetic status of individuals (Court-Brown 1967, Harris 1970), improved models of small, nonrandomly breeding populations (Karlin 1969), and long-term studies of primate populations including observations on obvious genetic malformations (Itani et al. 1963, pp. 29–36) and genealogy (Sade, Chapter 17).

Third, the study of primate diseases (Cockburn 1963, Fiennes 1967, Bray 1968) offers an excellent testing ground for theories of parasitism,

prudent predators (parasites), and the mutual adaptation of host and parasite. Mathematical models of the transmission of malaria, for example (Macdonald 1957), require primate studies when monkeys as well as men belong to the reservoir of potential hosts (Contacos and Collins 1969). The same conclusion applies to mathematical models of schistosomiasis (Goffman and Warren 1970) for the same reason (Miller 1960).

Several potentially applicable developments in peripheral areas related to primatology have already been mentioned. Others have been reviewed in the fields of social psychology (Abelson 1967, Whitla 1968), sociology (Coleman 1964a, 1964b), ecology (Watt 1968, Pielou 1969), political science (Bernd 1966), and stochastic models (Bartlett 1960, Bartholomew 1967). Recent flowering of the analysis of incomplete demographic data has immediate utility in primate studies, a field in which observations are often fragmentary (United Nations 1967).

What is required of data collectors and of data analyzers in order to promote full and rapid employment of available mathematical models, and development of new mathematical models, in primate behavior, sociology, and ecology?

As for the data generator, I subscribe to the dictum which E. O. Wilson enunciated at an informal Harvard seminar: the state of biology is such that a person at any level of mathematical sophistication can make substantial contributions. Nevertheless, the magnitude and effectiveness of those contributions can be improved by sufficient mathematical self-confidence to scan theoretical literature critically and to search for conclusions or predictions relevant to field work.

A simple example illustrates how an appreciation of the models which underlie field techniques can improve the use of field data. A familiar method of estimating the density of points in a plane is to choose random points and measure the distance from each to its nearest neighbor. The density of points may be estimated from the square of the mean of the distances or from the mean of the squares of the distances. Kendall and Moran (1963, p. 38) have shown that it is more efficient (i.e., gives an estimate with smaller variance) to use the latter estimate than the former. Hence following the inattentive routine of just presenting the average of nearest-neighbor measurements would not be taking full advantage of available information. Such mishandling of field data can destroy valuable information in them.

I would urge field primatologists to publish as full and detailed accounts of systematically collected data, along with the methods of collection, as possible. And I would urge them to encourage their students to have calculus through ordinary differential equations, probability theory, and some modern algebra. Properly taught, such courses can lay the foundation for confident consultation with active modelers.

The mathematical modeler also needs to gain an appreciation of what

is important and what is feasible in primate field studies. He can gain such appreciation by participating actively in primate field studies, or by consulting frequently and for a long time with field workers, or by trusting his armchair insight and luck. The first option is best. The armchair is an ideal vantage point from which to overlook the obvious and make impossible demands: when estimating density by choosing random points and measuring the distance to the nearest neighbor, the field worker must have alternative procedures if the random point is in a patch of six-foot-tall grass occupied by an elephant.

A warning issued to physicists by Bridgman (1927, p. 209) has equal relevance to those who construct and take seriously mathematical models in the life sciences:

> There is an aspect here of our physical research that is often lost sight of, namely, the small proportion of successful discoveries compared with the number of investigators. Certainly the number of unsuccessful attempts, even in the case of those fortunate individuals who make the great discoveries, is very much greater than the number of their successful attempts. (Faraday's reputed satisfaction with a 1/10% return comes to mind.) This must always be taken into account in estimating the probable chances of correctness of any new theory. With so many physicists working to devise new theories, the chances are high that many false theories will be found, in which a number of phenomena may apparently fit together into a new relation, but which eventually prove to be inconsistent with other phenomena, so that the proposed theory has to be abandoned. As physics advances and the number of investigators and the amount of physical material increases, one has to be more and more exacting in one's requirements of a new theory.