3

The Systems Dynamics Approach

The State Space

The *state space* of any system is simply the list of variables and their ranges describing the boundaries of the system.[1] The term *state* is used synonymously with the term *variable,* defining some characteristic of the phenomenon which takes a range of values. System states, like other variables, may be discrete (that is, "qualitative") or continuous (that is, "quantitative"). For example, the state space of a model describing political relations between two nations might contain the qualitative state "war," which is thought of as being either present or absent; an alternative formulation of the same problem might see the state space as containing a quantitative variable "level of aggression" that varies continuously from none to total war. In studying the dynamics of interaction among a small group of actors (individual persons, business firms, clans, etc.) one might conceive of the problem of the structure of network connections among the actors as one of qualitative states (i.e., each pair of actors are or are not "tied" to one another) or quantitative (i.e., the strength of the tie between each pair of actors varies in intensity from zero to some upper limit).

The system dynamics approach generally, and the DYNAMO language in particular, has a bias in favor of quantitative states. Because of the nature of the language it is easier to talk about continuously varying quantities than discrete states. Other languages for modeling continuous time dynamics are more balanced in this regard, but have other features that can create awkwardness of use.[2]

The systems dynamics tradition is somewhat unusual among systems approaches for the way that it conceptualizes and describes state spaces. All quantities describing the status of the system in question are regarded as being either "material" or "informational," with separate vocabularies and syntaxes being applied to the two categories.[3] Other approaches within the systems tradition do not draw this sharp distinction.[4] While the distinction between material and informational states is not necessary in the DYNAMO language, the structure of the language predisposes one to think in these terms.

"Material" and "informational" states are distinguished by writers in the systems dynamics tradition because it is felt that the principles governing the dynamics of physical quantities are fundamentally different from the dynamics of informational quantities.[5] Material quantities are regarded as "conserved" in that they persist over time and must occupy some state at all times. The most obvious examples of such things are physical objects such as people, money, and machines. These objects may occupy only one status at a time (e.g., each person is either young, middle aged, aged, or deceased). Another, less elegant way of putting it, is that material quantities are "used up" by being "used." For example, once the transition from youth to middle age has been made, individuals who fall in the middle aged category do not fall in the youth category (however much they may wish to).

"Information" states, in contrast, are regarded by systems dynamics as being fundamentally "nonconserved." That is, "information" is not "used up" by being "used." If I know something and tell you, we both know. This is a "nonconserved" dynamic. Material objects (such as money) may used by actors as "signals" to convey information, but are nonetheless conserved quantities. If I have a dollar and I give it to you, I no longer have it. This is a material or "conserved" dynamic. The distinction between kinds of states in the system dynamics tradition is both conceptual and mathematical. Quantities that persist—"material" states—are described by the mathematics of integration and the calculus, whereas quantities that are nonconserved—"information" states—are described by the algebra of differences.[6]

The distinction between material and informational states in the system dynamics approach is both attractive and troubling. At a philosophical level, critics both within and without the systems tradition reject the distinction as a false dichotomy. Information theorists, for example, insist on the treatment of all "states" as informational; the new physics is struggling with the seemingly no longer valid distinction between material and informational. At the mathematical level as well, there has been criticism of the distinction between "conserved" and "nonconserved" states. One consequence of rigidly maintaining the distinction is to lead to models that contain both differential and difference equations and hence draw on both calculus and algebra to describe state space dynamics. As the more mathematically inclined correctly point out, the distinction is unnecessary and inelegant from a mathematical point of view, and restricts the applicability of the tools of direct solution to systems dynamics models.[7]

The systems dynamics approach distinction between material and informational states can also be defended against these criticisms, at least on pragmatic grounds. While both the philosophical and mathematical critiques have a good deal of validity, the particular distinction between conserved and nonconserved quantities can be of considerable utility in social science construction.

At the philosophical and conceptual level, the distinction between material things that are persistent and "conserved" on one hand and "informational" things that are nonconserved on the other accords well with the way that most social scientists conceptualize social behavior. In thinking about social dynamics, most theorists tend to distinguish between acts and the meanings attached to acts. "Action" and "interaction" by themselves are most often seen as having a physical (behavioral) character that is logically separable from the meaning attached to them; action and interaction become "*social* action" and "*social* interaction" when they take on meanings as symbols for other actors. By distinguishing sharply between material and informational states, the systems dynamics approach and the DYNAMO language tend to structure the analyst's thinking about dynamics along those same lines most commonly used by sociologists, anthropologists, political scientists, and historians.

The mixed difference and differential mathematics peculiar to systems dynamics models restrict their analyzability by direct solution. At this point, there is a clear disjuncture of approach between that of mathematically inclined practitioners (especially in political science, economics, and sociology) and the intent of modelers in the systems dynamics tradition. Mathematical modelers dealing with continuous state/continuous time dynamics have shown a strong preference for models expressed as simultaneous linear differential equations. This preference is based on the use of direct solution as the method of choice for analysis of, and deductions from, the formalized theories. The systems dynamics approach is predicated on experimentation and simulation, rather than direct solution, as the primary method by which theorists can understand and make deductions from their theories. Simulation and experimentation are less powerful methods of understanding and analyzing a theory and its consequences than direct solution; however, greater flexibility of expression is obtained, and as a consequence far more complex phenomena and theories can be expressed. The mixed mathematics capabilities of the systems dynamics approach does not prevent the formulation of strict differential or strict difference equation models. It does lead one in the direction of flexible expression and away from mathematical analyzability.

The distinction between material states and informational states in the systems dynamics approach and the DYNAMO language is, therefore, consequential and reflects certain biases. By and large, these biases may be appealing to theorists interested in constructing formal theories of social dynamics. The distinctions between "conserved" and "nonconserved" quantities or between "material" and "informational" things seems to accord well with the distinctions between behaviors and meanings common in social science discourse. The greater ease of expression possible utilizing mixed mathematics and abandoning direct solution as a method of deduction is also appealing because it allows for the expression of quite complex relationships and for analysis by simulation, experimentation, and discovery, rather than by direct deduction and solution.

Material States: "Levels"

The most basic elements of theories expressed in the system dynamics language of DYNAMO are material states, called *levels.* In approaching the construction of a dynamic model, the identification of the levels of the system is the place to start. For example, in the simple model we examined above as Figure 2.1 and Figure 2.2, the levels of the system were the number of young, middle aged, aged, and deceased persons. These levels are continuous variables that are "conservative" and accumulate over time at rates governed by causal variables.[8]

In the diagraming conventions of the DYNAMO language, "levels" are represented as rectangles, with arrows flowing into and/or out of them.[9] The imagery is from fluid dynamics, and is intended to suggest a tank or storage location for quantities that flow into and out of the state, as in the flow of water into and/or out of a tank. There is no necessity in adopting the particular conventions of the DYNAMO language to represent the states and connections among them in a theory about dynamics. We will use these symbols throughout the volume, however. The more general languages of flow diagrams and circuit diagrams could also be used effectively to represent dynamics, but do not have quite the same evocative quality as the DYNAMO symbols, and are not as closely tied to the DYNAMO language.[10] At various times in this volume we will use both DYNAMO diagrams and simpler (but less specific) "circles and arrows" diagrams of connectivity, as suits the needs of the presentation. All theory building exercises usually begin with the simpler form of diagrams that show only the elements and connectivity among them; DYNAMO diagraming conventions are a useful tool in the step of translating such diagrams into equation form.

"Levels" are used in the DYNAMO language with a very specific syntax. The content and structure of the "sentences" are important in helping to structure thinking about dynamics. The statement below is a fairly typical "level equation" describing the "material state" of the number of persons in a population (POP.K). It illustrates all of the important features of the the systems dynamics language for describing the dynamics of such states.

```
L    POP.K = POP.J+DT(BIRTHS.JK+IMMIG.JK-DEATH.JK-EMIG.JK)
```

This somewhat intimidating-looking expression can be readily translated into plain English: The number of persons at time point "K" (POP.K) is equal to the number of persons at the previous time point "J" (POP.J), plus the integration or accumulation over time (DT) of a quantity. The quantity, in this case, includes the rate of births during the time interval between J and K (BIRTHS.JK), the rate of immigration (IMMIG.JK), the rate of deaths (DEATH.JK), and the rate of emigration (EMIG.JK). The "L" in at the beginning of the line is used to identify the "equation type," and is used by the DYNAMO simulation routines to control the order in which calculations are performed. There are a number of things to note about the way that this sentence is constructed.

First, note that the dynamic relation is conservative in the sense discussed previously. Population at a later point in time is equal to population at the earlier point in time, except as modified by births, deaths, immigration and emigration.[11]

Second, note that the "dependent variable" on the left side of the equation has a single time script: It describes the status of the variable at some particular instant in time (that is, "K"). Because this is the case, the level or status of the variable in the state space always refers to a quantity of things—such as numbers of people, percentages of national product, or degree of attitudinal support. It is a good practice, in thinking about the states of a system, to clearly define the "units" of all such variables as part of the process of defining terms.

Third, in contrast to the single time script of the "dependent" variable, each of the quantities in the expression to be integrated (that is, births, deaths, immigration, and emigration) carry two time referents: "JK." These quantities are called "rates" (more on them later), and are expressed in units per unit time. That is, for example, BIRTHS.JK in the level equation above are measured in numbers of events occurring between time point J and time point K. The "causal" factors determining population (that is, the independent variables in the equation, if one prefers), then, are explicitly dynamic quantities. They express the rates at which certain processes are occurring with respect to time.

The final important thing to note about the syntax of statements about "material states" is that there are numerous elements on the righthand side of the equation. Alternatively, one might say that there are several simultaneous causal processes producing change in the system state. In this particular case, two processes produce increments in the population (the rate of births and the rate of immigration), and two produce decrements (the rate of deaths and of emigration). Each of these processes is explicated in greater detail in "rate" and supporting equations that we will consider shortly.

The syntax of the level statement with its multiple possible causal processes is a good stimulus to clear thinking about the dynamics of states. The structure of the statement itself leads one to ask what causes the level to go up or down, and do each of these causal processes have the same or separate determinants? The syntax should lead one to consider whether the causes for increases and decreases are the same. In the case of population dynamics, of course, they are not; many social processes, however, may be reversible in the sense of having the same causes for both increases and decreases in level over time.

The syntax of the level statement also urges one to think in multivariate terms by allowing the easy expression of multiple simultaneously operating causal processes. This is the "normal" way of thinking about problems in multivariate statistical models, but is not always the language of verbal theory. Verbal formulations tend to be highly simplified (and perhaps oversimplified) ways of stating theories as bivariate "propositions"; the systems dynamics level equation leads one to automatically consider the simultaneous operation of multiple causal processes. The syntax of the level statement makes quite explicit what most social scientists mean when they refer to changes in the levels of particular variables over time as resulting from the simultaneous operation of many factors or the "conjuncture" of historical forces.

Occasionally it is useful to think about a level that has only things "flowing into" it, or only "flowing out" of it. In the simple population dynamic that we examined as Figure 2.1 and 2.2, for example, the level "deceased" is such a quantity. Individuals flow into this level, but they do not exit (at least for the purposes of most social science models). In the language of systems, states that have only processes incrementing them over time are called *absorbing states,* while states that have only processes decrementing them over time are called *sources.* In the peculiar jargon of the language of the system dynamics tradition, these special kinds of levels are termed *sources* and *sinks* and are represented in diagrams with a special symbol resembling an amoeba.

While it is perfectly possible to formulate a dynamic model without giving special consideration to the question of where material quantities ultimately come from and where they finally go, the existence of the special symbols and concepts of sources and sinks can lead one to ask important questions that help to clarify and specify theories. Too often we do not realize that we are theorizing about a part of a process in our theories, not the whole phenomena from source to sink. These devices, of course, also lead to asking silly questions that result in trivial answers, as in the simple population model example.

Most of the dynamic processes addressed by social scientists tend to involve relatively simple chains of levels governed by quite complex control systems. The notions of sources and sinks are helpful in improving theory specification about such processes in two ways. First, they lead one to consider, for each state in the theory, what the previous link in the causal chain was, and what the next link in the causal chain may be. This can often lead to elaboration of the model in interesting and valuable ways. Second, the specification of the sources and sinks of "conserved" processes provides one way of understanding the boundaries of the phenomenon analyzed by the theory. We may choose, for some purposes, to regard the source of a material state as exogenous. In representing this as a diagram or level equation, the exogenous variable becomes a "source." Designation of a level as a source indicates that we are not going to specify its causes as part of the theory (i.e., it is "exogeneous"). Similarly, designation of a given level as a sink indicates that we regard it as having no consequences for other variables in the theory. Both types of statements are thus clear ways of identifying some of the "limits" of the theory we are developing. As with all statements of limits, the choice of which states are sources and sinks is a pragmatic one, defining the boundaries of the phenomenon for the purposes of the construction of the theory.

Specification of the material levels or states of a system is the first very necessary step in fully formalizing a dynamic system. Once we understand the boundaries and limits of the material or conserved quantities, the next step is to describe the informational levels and connections among them that complete the definition of the state space of the theory.

Information States: "Auxiliaries"

A large proportion of the state space of most models involves descriptions of the current levels of "information" or "nonconserved" quantities. For example, the perceptions of acts, cognitions about them, formulation of decisions, and mapping of strategy can all be considered "informational." Economic actors make decisions on the basis of prices, political actors seek to gauge "public opinion," individuals form attitudes on the basis of observing behaviors. All of these processes involve "levels" of information. More generally, all of the elements of the "control" processes that govern the dynamics of change in levels or material chains are represented in systems analysis generally (and system dynamics particularly) as informational.

In our simple model of population dynamics (Figure 2.2) for example, the "informational" or "control structure" is the portion of model that shows how the rates of "flow" between the "levels" (i.e., young, middle aged, aged, deceased) are controlled by information flows.

Because of the special importance assigned to "information" in systems theory, a somewhat different vocabulary and syntax has been developed by systems dynamicists than that normally used for "material" states. In the diagraming conventions of DYNAMO, informational states are represented as circles, with "flows" of information denoted by dashed lines. To serve as a starting point for discussion, part of our original population dynamics model (Figure 2.2) has been elaborated with these symbols in Figure 3.1.

The informational level "number at risk" in this very simple example is a direct function of the material level "number of young." The dotted line connecting the "number of young" and the "number at risk," however, does not represent the actual movement of people from one state to another. Rather, the flows of information are "nonconservative," and the dashed line indicates only the "take off" or "monitoring" of information about the material state "young," not the actual movement of persons. Similarly, the connections between the informational state "number at risk" and the flows between the states of young and middle aged, or the states of young and deceased, do not represent movement or change in physical things. The "number at risk" is a piece of information that is used in the determination of the rates of transition between states, but the information is not "used up" by being used. Hence, this "flow" as well is represented with a broken line and the little circle representing an information "take off" or "monitoring."

There are also, in Figure 3.1, two "information states" represented by the special symbols of circles with an intersecting line segment. These "information states" are "constants"—quantities that are used in explicating the control system, but which are not, in themselves, determined by other variables in the theory. These quantities are the

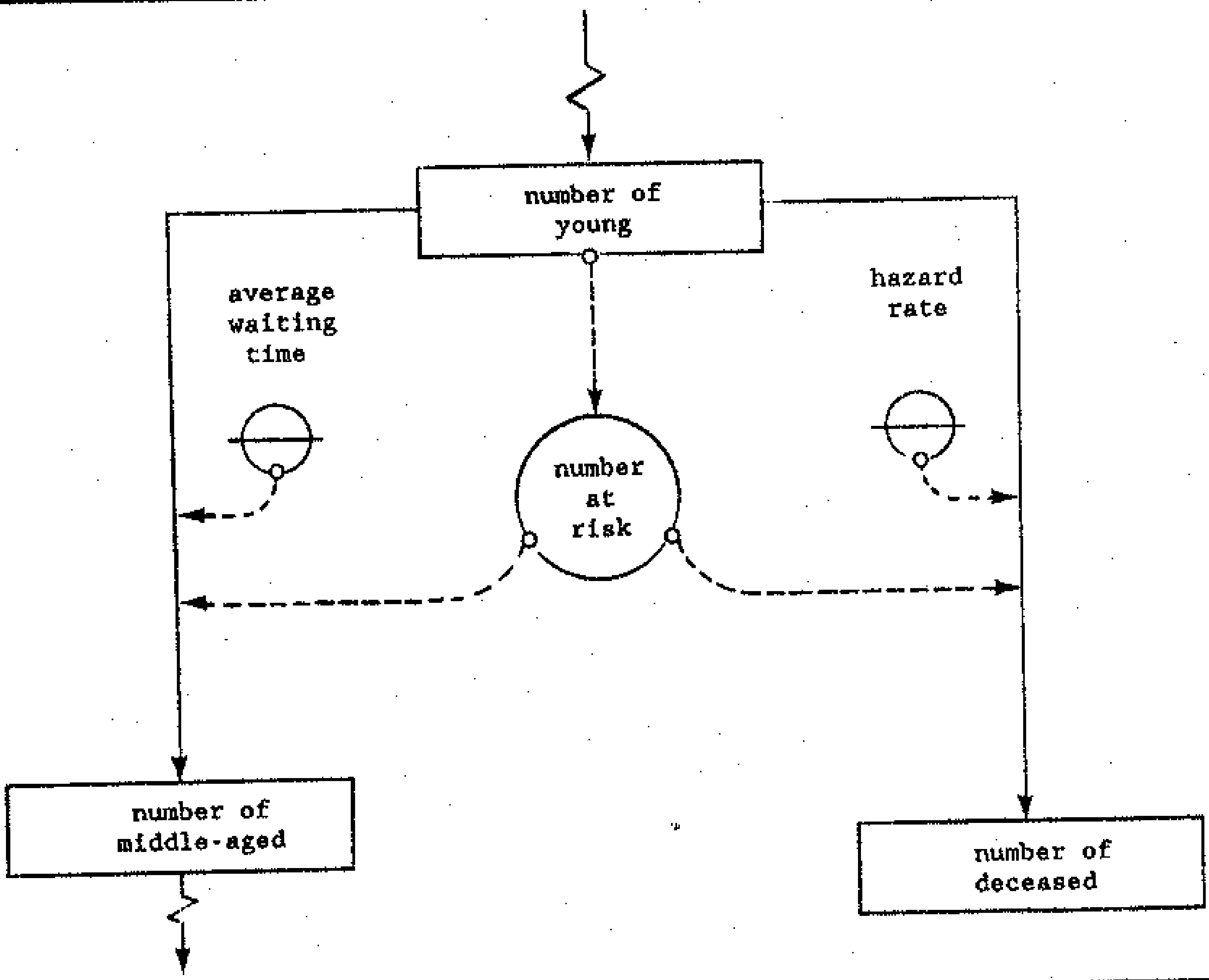

Figure 3.1: The roles of auxiliaries and constants.

informational equivalents of the "sources" and "sinks" of the conserved parts of the state space. Of course, in our simple model, a next step of elaboration might well be to turn one of the "constants"—the hazard rate—into a "variable," or information state that is a function of other variables.

The information states, or "auxiliaries," of most models describing social dynamics can be numerous and connected in complicated ways. Suppose, for example, that we were modeling the interaction between two nations engaged in an arms race. (We will develop this particular model at some length in a later chapter.) Each nation bases its behavior on its own goals, preferences, and capacities, but also on its perceptions of actions being taken by the other. In the simplest case, these perceptions might be entirely accurate and nonproblematic monitoring of the level of armaments held by the opponent. This kind of relationship would be expressed in the sentence (equation) in DYNAMO:

```
A    PARMS.K = ARMS.K
```

That is, the level of arms perceived (PARMS.K) at time K is equal to the actual number of arms at time K (ARMS.K). Note two things about this statement: It is not "conservative" (at each time point, PARMS is recalculated from other quantities; there is no PARMS.J in the equation for PARM.K). And the creation of the information state PARMS.K does not result in the destruction or transformation of ARMS.K. That is, the creation of information by perception does not use up the object perceived.

It might well be that the informational monitoring process is not so simple as the auxiliary equation above proposes. Lets look at an alternative specification of the process that creates the informational level "perceived arms" or PARMS:

```
A    PARMS.K = DELAY3(ARMSI.JK,2)
X    +CONST(ARMS.K+NORMRN(FEAR,10))
```

Because of its length, it was necessary to carry this statement beyond a single "logical record," and the statement is continued on a second (X-type) record. This formulation suggests a much more complex process of how information about the level of an opponent's arms is perceived. Roughly, this statement can be translated as follows: The level of arms that we perceive our opponent to have at time K (PARMS.K) is the sum of two complex quantities. The first is a delayed perception (in this case, a third-order exponential delay with an average length of two time periods: DELAY3(ARMSI.JK,2)) of the rate at which our opponent has been adding to his arms stock over the preceding time period (ARMSI.JK). The second quantity is some fixed multiplier (CONST) of our opponent's current level of arms (ARMS.K) plus an amount of normally distributed noise (NORMRN). The amount of this noise, or error in perception, has a mean of some constant quantity (FEAR), and a standard deviation of 10 units.

This particular specification is probably not a very useful one for describing the cognitive processes that actually occur in arms races. It does, however, illustrate some important points. By distinguishing clearly between "material" flows on one hand, and the "information" that is monitored and transformed to "control" these flows, a great deal of structure is forced on the theory constructor. In this simple example we are led to the important insight that arms races are governed by perceptions of threat that may be imperfect reflections of the material conditions. And we are required to be quite specific about where information comes from, how it is transformed, and where it goes to create the control system that governs the flows and transformations of the material states of the system.

The syntax of statements about the determinants of informational levels ("auxiliary" statements) is very flexible.[12] Virtually any kind of terms may be used, and they may be combined in virtually every possible fashion, including special functions involving time and nonlinear relations. Informational levels can be created out of information "monitored" from material states, other informational states, rates of change, and constant terms. In our example above, a variety of constants (FEAR), levels (ARMS.K), and rates of change in levels (ARMSI.JK) are all used to create the perceived level of arms. None of these quantities on the right-hand side, however, are "used up" or transformed in the process.

The "independent variables" in the auxiliary equation can be combined in a variety of ways. Often simple linear relations are plausible specifications (addition, subtraction, multiplication, division), but often nonlinear combinations (such as thresholds), or combinations involving noise, delay, and distortion are necessary to mimic the informational processes of social actors. A variety of shorthand tools for some of the most common types of complex relations are provided in the vocabulary of the language itself (see Chapter 4), and others can be created (user-defined macros). The much greater flexibility of the language when dealing with information than with material things reflects the prior assumption that information dynamics are fundamentally different from material dynamics.

The state space of any pattern of social dynamics, then, can be defined using the "level" and "auxiliary" equations of the DYNAMO language. The particular syntax of these forms of statements about the elements of the state space reflect some peculiarities of the conceptual approach of a particular theoretical school—that of "systems dynamics." These peculiarities of the language for describing the state space of a theory are, in most cases, helpful and consistent with the way that most theorists think about social dynamics. The language, however, is also quite flexible (though not infinitely so), so that questions of what are "levels" and "auxiliaries," as well as what is meant by terms such as "sources," "sinks," and "constants" are open to the pragmatic definition of the theory constructor.

Once the elements of the state space, both material and informational, have been described, the next step in the process of theory building is to describe the processes that determine the rates of change in elements of the state space. That is, we must next specify how the causal connections work across states and time.

Rates of Change

The "levels" of a system define the status of its material states at a point in time, but do not speak directly to the issue of the dynamics of the system. "Level" equations make reference to rates of change (the "JK" terms), but do not explain these rates. "Auxiliary" equations, on the other hand, describe both the informational elements of the state space and the over-time relations among these elements. After the state space of a theory has been defined with the tools discussed above, attention must be turned to the dynamics of the conserved states: That is, what causes the state of the system to change from one time point to the next. In the system dynamics approach, the hypotheses about the causes of change are embodied in separate statements, called "rates," with their own special syntax.[13]

Recall for a moment our earlier example of a level equation describing a very simple model of population dynamics:

```
L     POP.K = POP.J+DT(BIRTHS.JK+IMMIG.JK-DEATHS.JK-EMIG.JK)
```

Again, this statement says that the size of the population at instant K is equal to the size of the population at some prior time point, J, plus the integration (or accumulation, if you prefer) of births, deaths, immigrations, and emigrations that occur at certain rates per unit time across the interval between J and K. Births, deaths, immigrations, and emigrations are thought of as occurring continuously across the time interval between J and K.

The purpose of rate equations is to describe the "causes" of these rates, that is, to specify the effects of causal factors on the rate at which the level (in this case, population) is changing. Causal effects operate on the rates of change, which are decomposed into a series of separate processes: births, deaths, immigrations, and emigrations. The level of the population, then, is really just a momentary snapshot of the accumulated consequences of causal processes (rates) that are occurring continuously in time.

Systems dynamics models, like models formulated more directly into differential equations, are really speaking to the causes of rates of change in the states of the system. A system dynamics flow diagram of our simple population model will help at this point both to reiterate the notion of dynamic models as revolving around rates of change and to introduce diagraming conventions.

The symbol for a "rate" in the systems dynamics flowgraph conventions is "milk-can" shaped and is intended to invoke an image of a

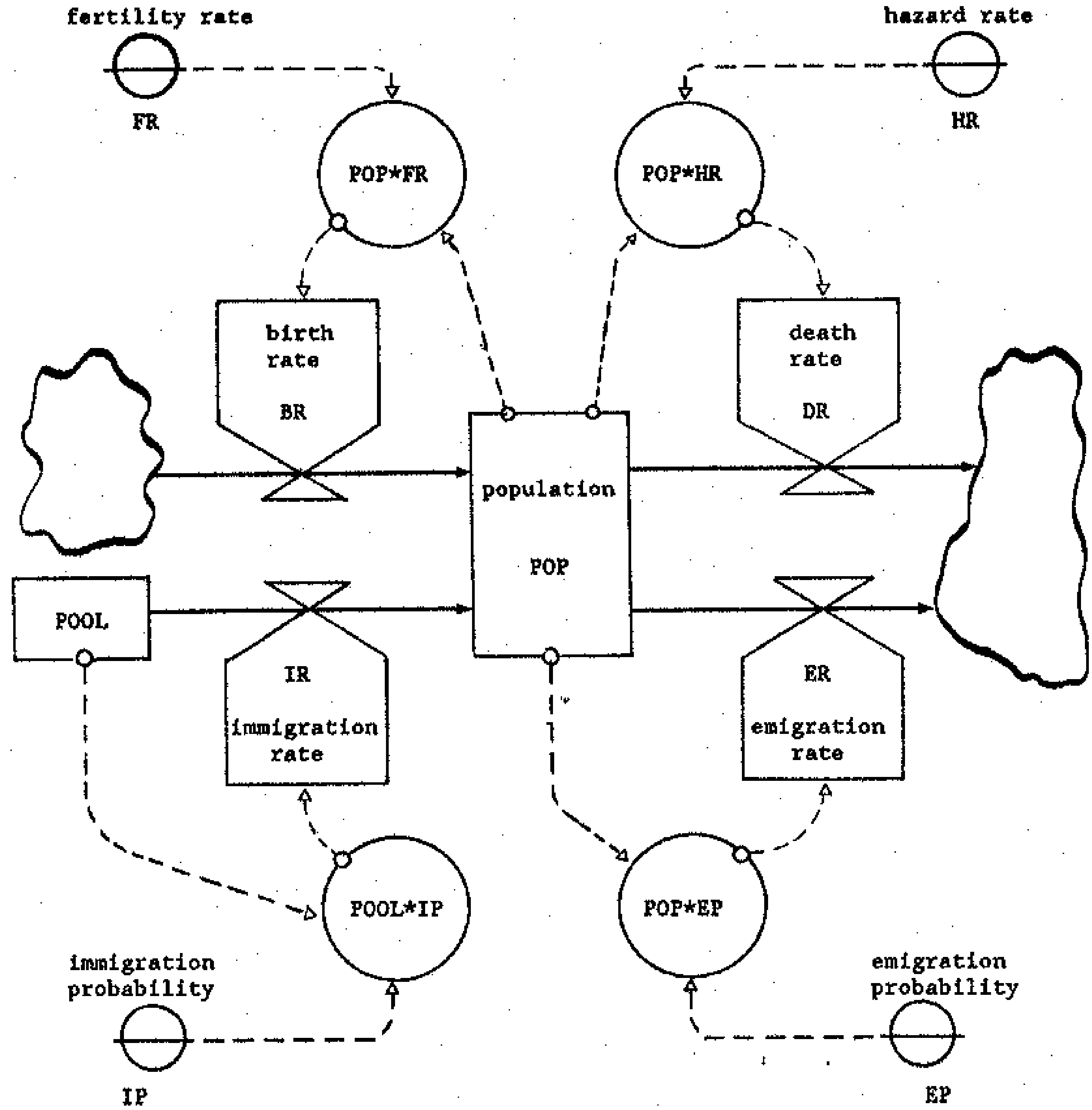

Figure 3.2: Rates in a simple population model.

gateway or valve controlling the "flow" into or out of a level. All flows into and out of levels must contain such valves to define the rate at which increments or decrements to the level in question occur. In our current model, population change is a result of four forces that either add to or subtract from the level of the population and which have different determinants. If any of the flows in question (that is, any of the rates) had the same causes, they could be combined into a single-rate statement, though the theorist might wish to keep separate statements for each theoretically important factor for ease of explication. Again, mathematical elegance is not the chief goal of formal theory in the DYNAMO approach.

The diagram is very useful in showing what factors are hypothesized to contribute to the rate of change in population by their effects on births, deaths, immigration and emigration. In our simple model, we show the simplest possible specification: Each rate (BR, DR, IR, ER) is determined by the current level of population (that is, the system is, in one sense "self-referencing") and by a constant. The diagram, however, only shows which terms are connected to which rates of change and is usually not sufficiently for complete specification. The statement of the determinants of rates of change in equation form is used to make these specifications. For the model in Figure 3.2, the "rate" equations are shown below.

```
R    BR.KL = POP.K*FR
R    DR.KL = POP.K*HR
R    IR.KL = POOL.K*IP
R    ER.KL = POP.K*EP
```

The first statement says, rather obviously, that the number of births occurring (BR.KL) between time point K and time point L (K and L are used, rather than J and K, to control the order in which calculations are done when simulations are performed) is equal to the number of persons in the population times a constant (FR). The fertility rate (FR) in this case is the probability that a given person will give birth in the interval between K and L. The other rate equations (for DR, IR, and ER) are similar, but use different constants: a "hazard rate" (probability of death over the interval from K to L); an "immigration probability" (the probability that a given member of the pool of possible immigrants actually immigrates between time points K and L); and, an "emigration probability" (the probability that a member of the focal population will emigrate in the time period between K and L).

"Rate" equations are largely unrestricted in form, and they can grow to be quite complex. One can imagine, for the current simple population model, that one might wish to replace the constants (FR, DR, IR, and ER) with more elaborate expressions. The fertility rate, for example, might be made a variable by making it a function (described in an "auxiliary" equation) of the average age of the population, levels of economic performance, values about desirable family size, etc. These quantities, in turn, might be specified to be functions of other levels, auxiliaries, and rates. In principle, all of this complexity would properly be a part of the "rate equation" or sentence. In practice, since the determinants of rates are often hypothesized to be extremely complex and involve many other variables, "auxiliary" type equations are used to express the portions of the rate. This can sometimes lead to confusion, as "auxiliaries" are used in the DYNAMO language to describe both "levels" of "information" and to perform the auxiliary calculations that enter the right-hand side of rate equations.[14]

The "rate" equations are the focal point for hypotheses about the dynamics of systems. In systems dynamics models (and differential equation models generally), the status of the system at any point in time (that is, the state space) is the result of continuously ongoing causal processes (that is, the rates). Hypotheses about the determinants of variation in the rates of change in dynamic models are, consequently, where the real "causal" theorizing occurs.

The syntax of the DYNAMO language is of assistance in the effort to think clearly about the causes of change in social phenomena. For each "level" in the system, one or more "rates" must be defined. By considering the specific "flows" that increment and/or decrement the level, the problem of specifying the causes of change can often be divided into several simpler (but simultaneously operating) processes. For each of these causal processes (i.e., rates), the language requires that we identify the factors that explain the rate in question. That is, the specification of the rate equation requires that we think clearly about what other levels, rates, auxiliaries, and constants determine the rate. This is equivalent to identifying the "independent variables" that have dynamic causal impacts. Lastly, the language requires that we write a specific hypothesis about how these factors "fit together" to produce change in the rates.

Because the language of rates and auxiliaries is very flexible, virtually any form of relation among independent variables and the "rate" can be specified. It is often best to work backwards by first asking what determines the rate and then what determines these factors, and so on until the "chain" of expressions is complete.

For example, let's look at the birth rate in the example that we've been considering. The birth rate itself is determined by a very simple expression: The number of births between two points in time is equal to the population "at risk" times the probability of an event—that is, the fertility "rate" or probability. The size of the population at any point in time is not problematic; it is accounted for by the level equation. This part of the "chain" is complete, as it connects directly to an already fully specified part of the model (the level of population). The causes of the fertility rate, however, could be further elaborated. We might wish to formulate a theory of fertility probability, perhaps based on the age structure of the population and the economic wealth of the population. This would require us to calculate these quantities, or to supply them as constants (that is, to specify the age structure and wealth levels as the "left-hand" side of either auxiliary or constants types of equations).

Again the theorist faces a forced choice: To treat age structure and economic wealth as exogenous, or to provide expressions that specify their causes.

This method of working backwards in the elaboration of the determinants of rates creates models that are fully self-referencing (that is, "closed systems") or have specific points at which simplifying assumptions have been made. In Figure 3.2, for example, we have as much as stated that our "theory" of population dynamics will not deal with the causes of fertility, mortality, immigration, and emigration probabilities; instead these factors are treated as constants. Several things are accomplished by this method. First, of course, we become quite clear about the boundaries of our theory. Wherever constants enter the chains that describe the causes of rates of change, limiting assumptions are being made. Secondly, the method of working backwards helps us to take apart what are often very complex causal processes into a series of simpler, sequential processes. It is not uncommon to have as many as 10 or 20 "auxiliary" and "constant" types of statements to specify the process that determines a particular rate. Mathematically, this is extremely inelegant; but the clarity of thinking and clarity of expression gained in describing the (often complicated) processes that cause change is usually worth the price.

Conclusions

The systems dynamics approach to formalizing theories about dynamics is one of several traditions within "systems theory" and "systems analysis." The conceptual approach of system dynamics and much of the method of constructing theory from this perspective is embodied in a particular formal language called DYNAMO. In this chapter we have examined generally how the "systems" analysis tradition thinks about social dynamics and, more particularly, have examined the basic elements of the "systems dynamics" approach.

The "systems dynamics" school of theorizing about dynamics has provided one such "higher" language for constructing theories about continuous state continuous time dynamics: DYNAMO. This particular language for describing state spaces and the dynamic connections among the elements of state spaces has a number of conceptual peculiarities, unique vocabulary, and uncommon syntax. The structure of this particular language, however, is unusually useful for formalizing social scientists' theories.

Systems dynamics distinguishes between "conserved" and "nonconserved" elements of the state space of a theory, and provides different vocabulary and syntax for describing these types of elements (e.g., "levels" and "auxiliaries"). This rather unusual distinction reflects both a conceptual and mathematical approach. The language can be used with or without its notions of differences between "material" and "informational" states, and can be used with either differential mathematics or difference mathematics, or both. The distinctions among types of states, however, is often an aid to thinking about social science problems where behaviors and meanings are regarded as separate, but connected phenomena. The use of both differential and difference types of expressions allow for great flexibility of expression, a desirable feature to most theorists.

The systems dynamics language also has a somewhat peculiar approach to dealing with the formalization of theory about the dynamic connections among states. Where differential equations are written in unified form, expressing the impacts of independent variables on the rates of change in dependent variables, the DYNAMO approach takes such expressions apart into (often large numbers of) "level," "rate," and "auxiliary" equations. These equations are mathematically inelegant, but allow the theorist to reason through the causal chains that determine the rates of change in states. While the statements are not compact, each component expression tends to express a rather simple (and hence, comprehensible) part of the overall causal relation, and the connections among the parts of the causal statements can be easily traced.

The DYNAMO language, like all other specialized languages for accomplishing particular expressive tasks, provides some shorthand tools for accomplishing complicated and often repeated operations. In formalizing theories about the dynamic relations among the elements of complex state spaces we don't need 110 names for different kinds of snow, but our task is considerably simplified if we have some shorthand ways of describing dynamic connections. The tool kit provides some of the most useful such "words".

Notes

1. More complete discutions of the idea of a "state space" may be found in Brunner and Brewer (1971), Chorafas (1965), Hall (1962), and Weiner (1948).

2. Mixed discrete and continuous state languages such as SLAM, GASP, and SMOOTH tend to be rather complicated, difficult to learn to use, and require substantial computer power. Other continuous state languages such as CSSL are elegant and efficient, but closely tied to the mathematics of differential equations. Languages intended primarily for discrete state models (e.g., SIMULA)

generally cannot be applied where there are any continuous states in the model. For some interesting discussions of the strengths and weaknesses of particular languages, see particularly Robinson (1972) and Buxton (1968). These discussions, while excellent, are considerably dated, as new languages appear quite frequently.

3. Many system dynamics approaches differentiate kinds of elements of the state space still further. Information is often divided into "data" and "orders"; material flows are often differentiated into "personnel," "money," "raw materials" and "capital." These categories are not as "primitive" as the information/material distinction, but are particularly useful for modeling management problems. See particularly Goodman (1974), Pugh (1980), and Roberts et al. (1983).

4. "Semiotics" and "information theory" tend to reduce all elements of the state space to cultural representations, or information. Many researchers in artificial intelligence, machine systems, and general system theory draw no fundamental distinction between types of state space elements—regarding each as a function or assembly process. See particularly Gilbert (1966) on information theory approaches.

5. Perhaps the best arguments in favor of this approach come from the founder of the systems dynamics school; see Forrester (1961, 1968).

6. In fact, the DYNAMO language is highly flexible, and need not be used in accordance with the philosophical strictures of the "systems dynamics" school. Conserved quantities may be treated with either algebra or calculus, as may nonconserved quantities. The common practice of distinguishing types of things and types of mathematics, however, is a generally useful thinking tool.

7. For a particularly insightful discussion of the peculiarities of the systems dynamics approach, see Day (1974). Pritsker (1974) provides an illustration of translation between DYNAMO and the more explicitly mathematical language GASP. Robinson (1972) also compares several languages approaches to the same problem.

8. There are a number of excellent introductory treatments of the details of the DYNAMO language and its implementation in simulation analysis. It is not our intent to duplicate these primers and users manuals in this volume. Those readers who wish to become proficient in the DYNAMO language will have to study the following sources: Goodman (1974), Pugh (1980), Pugh-Roberts Associates (1982), Roberts et al. (1983).

9. On the flow-diagraming conventions of DYNAMO, see the sources listed in the previous note, particularly Goodman (1974).

10. See, for example, work on path diagrams (a particular variant of linear systems) in Duncan (1975) and Blalock (1961, 1971). The more general diagraming conventions of "flow diagrams" (for continuous state systems) are examined in Heise 1975, Lorens (1964), and Stinchcombe (intuitively, at least, 1968). Methods for diagraming and "solving" discrete state systems by means of directed graphs are dealt with in a large number of sources, including Barnes (1972), Berge (1962), Busacker and Saaty (1975), Flament (1963), Harary (1969), Harary et al. (1965), Hoivik and Gleditsch (1975), and Huggins and Entwisle (1968).

11. "Level" equations need not necessarily involve integration of rates, though most continuous and conserved quantities are usefully regarded in this fashion. Nor do level equations have to involve conservation. That is, it is not necessary that the previous level of the system enter into the calculation of the current level of the system.

12. Auxiliary statements in DYNAMO may use virtually any functional form to describe relations among quantities and over time. Certain common functions, such as discrete lags, are more difficult than others to create in the language. For much greater detail, see the sources cited in note 8, above.

13. Differential equation models can be thought of as being composed of statements of the form: $dY/dt = f(M,N,P)$. Level equations perform the task of integrating Y, while the rate and auxiliary equations are used to specify the functional forms of the effects of M, N, and P.

14. This peculiarity represents a conceptual flaw in the DYNAMO language. The "states" and "rates" of material quantities are clearly distinguished by the use of separate equation types; the "states" and "rates" of informational quantities are both represented by the single "auxiliary" equation type.