Chapter 2

Simulation as a method

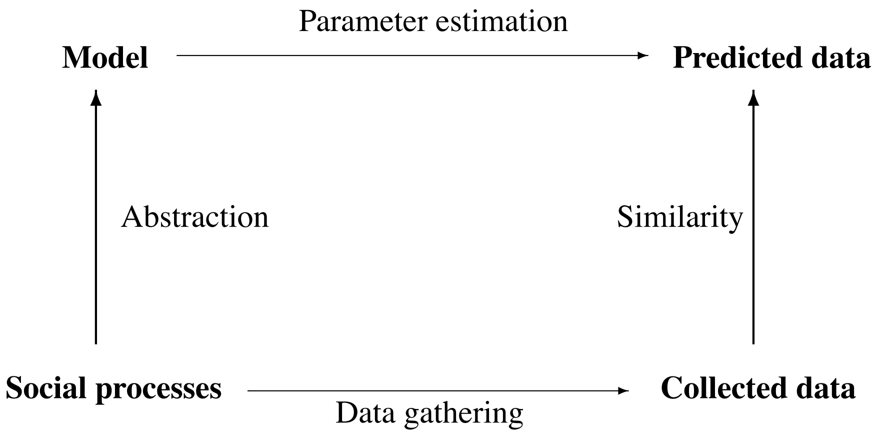
This chapter is about the use of computer simulation as a method of social research: the logic behind the method, the stages that one needs to go through and the pitfalls to be avoided. To start, we need to define some terms.

We shall assume that there is some 'real world' phenomenon which you, the researcher, are interested in. This we call the *target* (Doran and Gilbert 1994; Zeigler 1985). The aim is to create a *model* of this target that is simpler to study than the target itself. We hope that conclusions drawn about the model will also apply to the target because the two are sufficiently similar. However, since our modelling abilities are limited, the model will always be simpler than the target. For example, we might model the real market for the wholesale supply of fish with a simpler system where both suppliers and purchasers are represented by computer programs standing in for complex and multifaceted businesses and their customers (cf. Weisbuch *et al*. 1997).

In the social sciences, the target is always a dynamic entity, changing over time and reacting to its environment. It has both structure and behaviour. This means that the model must also be dynamic. We can represent the model itself as a *specification* – a mathematical equation, a logical statement or a computer program – but to learn something from the specification, we need to examine how the behaviour of the model develops over time. One way of doing this is using an analytical method. This entails deriving the model's future structure from the specification by reasoning, perhaps using logic or more often by using mathematics. For example, we might have a model of the relationships between a set of macroeconomic variables and use algebra to derive the outcome if one of those variables changes over time.

With complex models, especially if the specification is nonlinear, such analytical reasoning can be very difficult or impossible. In these cases, simulation is often the only way. Simulation means ‘running’ the model forward through (simulated) time and watching what happens. Whether one uses an analytical technique or simulation, the *initial conditions*, that is, the state in which the model starts, are always important. Often, the dynamics are very different depending on the precise initial conditions used.

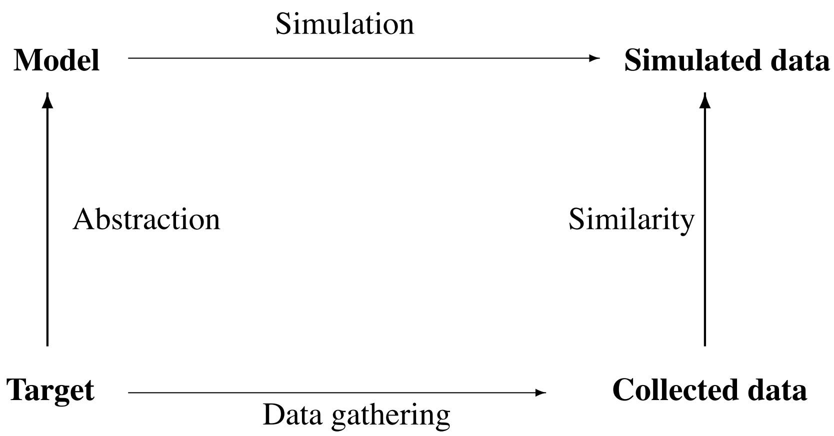
Figure 2.1: The logic of statistical modelling as a method (after Gilbert 1993)



The logic of simulation

With statistical models, the relationship between model and target is quite well understood (see, for example, Gilbert 1993: Chapter 1). As Figure 2.1 indicates, the researcher develops a model (for example, a set of equations) through abstraction from the presumed social processes in the target. These equations will include parameters (for example, beta coefficients) whose magnitudes are determined in the course of estimating the equations (this is the step where a statistical package would normally be used). As well as developing a model, the researcher will have collected some data with which to perform the estimation (for example, survey data on the variables included in the equation). The analysis consists of two steps: first, the researcher asks whether the model generates predictions that have some similarity to the data that have actually been collected (this is typically assessed by means of tests of statistical hypotheses); and second, the researcher measures the magnitude of the parameters (and perhaps compares their relative size, in order to identify the most important).

Figure 2.2: The logic of simulation as a method



Much the same logic underlies the use of simulation models, as Figure 2.2 shows. Once again, the researcher develops a model based on presumed social processes. But this time, the model might be in the form of a computer program rather than a statistical equation. This model is run and its behaviour measured. In effect, the model is used to generate the simulated data. These simulated data can then be compared with data collected in the usual ways to check whether the model generates outcomes which are similar to those produced by the actual processes operating in the social world.

Both simulation models and statistical models can be used for explanation and prediction. The prime purpose may be to try to understand some particular social phenomenon: for instance, why one gets clusters of people all of whom share the same opinions (see Chapter 7). We may not be particularly interested in predicting how many people there are in a cluster and, indeed, the theory might suggest that forecasting the number is impossible (see the discussion of complexity in Chapter 1). A model that incorporates simulated processes which lead to clustering might suit the purpose of explaining the observed clusters. In other circumstances, we might be particularly concerned with making specific predictions and less concerned with understanding social processes. For example, there might be a requirement to predict the level of aggregate personal taxation flowing to the state in 20 years' time, taking account of demographic changes (see Chapter 4). The model we construct might have little in it about the social processes involved in defining taxation policy and only very simple assumptions about demographic trends, yet be useful in making predictions about aggregate fiscal changes.

While some statistical and simulation modellers emphasize the desire for understanding and others emphasize the need for making predictions, all simulations have in fact to satisfy both requirements: a successful predictive model will contribute to understanding at least to some degree, while an explanatory model will always be capable of making some predictions, even if they are not very precise.

While there are these strong similarities between statistical models and simulation models (and the boundary line between the two is not a hard-and-fast one), there are also important differences. As we noted in Chapter 1, simulation models are concerned with processes, whereas statistical models typically aim to explain correlations between variables measured at one single point in time. We would expect a simulation model to include explicit representations of the processes which are thought to be at work in the social world. In contrast, a statistical model will reproduce the pattern of correlations among measured variables, but rarely will it be modelling the mechanisms that underlie these relationships.

The stages of simulation-based research

With these basic ideas about the logic of simulation in mind, we can outline the 'ideal' set of steps in using simulation in the social sciences (cf. Doran 1997b). One starts by identifying a 'puzzle', a question whose answer is not known and which it will be the aim of the research to resolve. For example, we might be curious about the reasons for the pattern of Puebloan settlements which were established in Mexico from AD 900 to 1300 (Kohler *et al*. 1996). This leads us to the *definition* of the target for modelling (settlement dynamics in the Mesa Verde region). Normally, some *observations* of the target will be required in order to provide the parameters and initial conditions for our model. For the work of Kohler *et al*. (1996), these were obtained from detailed archaeological work by Van West (1994). One can then make some *assumptions* and design the model, probably in the form of a computer program. The simulation itself is performed by executing this program and the output of the simulation is recorded.

So far, the steps involved are fairly obvious, although often not simple to carry out. The remaining steps often receive less attention, yet they are crucial. We need to ensure that the model is correctly implemented and working as intended. This is *verification* – in effect, a 'debugging' step. Unfortunately, this process can be difficult to carry out with complex simulations and, in particular, it is difficult to know whether one has eradicated all the remaining bugs. The difficulty is compounded by the fact that most social science simulations are dependent on pseudo-random numbers to simulate the effects of unmeasured variables and random effects (Gilbert 1996) and so repeated runs can be expected to produce different outcomes.

Next, there is *validation*, ensuring that the behaviour of the model does correspond to the behaviour of the target. If settlement patterns in the Mesa Verde are being modelled, the simulation needs to reproduce to some adequate degree the observed pattern of settlements. Unless there is some correspondence, the simulation is unlikely to be a plausible model of the processes which led to the formation of those settlements. Finally, one needs to know how sensitive the model is to slight changes in the parameters and initial conditions: *sensitivity analysis*. In the following we shall consider some of these steps in more detail.

Designing a model

Every model will be a simplification – sometimes a drastic simplification – of the target to be modelled. The most difficult step in designing a model is to decide what needs to be left out and what needs to be included. The more that is left out, the greater the conceptual leap required between the conclusions drawn from the model and their interpretation in relation to the target. The more that is put in, the more precisely the parameters have to be measured or assumed, and each of them may have an effect on the validity of the conclusions which are obtained. What one hopes for is a model that embodies the minimum number of assumptions, but which applies as generally as possible to many different circumstances. The choice of where to place one's model on this continuum between the detailed and the abstract is partly a matter of skill and experience, partly a matter of research style and partly a matter of the amount of data one has available and how difficult it is to collect more. In general, accuracy (in terms of the number of data points and assumptions built into the model) is important when the aim is prediction, whereas simplicity is an advantage if the aim is understanding (Axelrod 1997a).

The temptation is to make a model more detailed than it really needs to be. Apart from the sheer labour of collecting and entering what can quickly amount to many thousands of data points, there is a danger that the additional complexity of dealing with substantial quantities of data will mean that the stages of verification and validity become very difficult to carry out. This in turn means that valid conclusions will be hard to draw from the research. The best map of the world is the world itself, but unfortunately such verisimilitude teaches us nothing about how the world works.

At the other end of the continuum from detailed to abstract modelling is research on 'artificial societies'. This is simulation without reference to any specific 'real world' target. The object of study is the set of possible social worlds, of which the actual world in which we live is just one (Conte and Gilbert 1995). As Epstein and Axtell (1996: 4) write:

> We view artificial societies as *laboratories*, where we attempt to 'grow' certain social structures in the computer – or *in silico* – the aim being to discover fundamental local or micro mechanisms that are sufficient to generate the macroscopic social structures and collective behaviours of interest.

At the heart of research on artificial societies is the goal of finding theories that apply not just to human societies but to societies of interacting agents generally. For example, there are results about the consequences of constraints on communication in markets in which there are some agents selling and others buying (see, for example, Alvin and Foley 1992). These apply regardless of whether the buyers and sellers are people, organizations or computers. Another example of the value of experimenting with artificial societies is Doran's (1997a) work on foreknowledge. His simulation studies the implications of agents having knowledge of future facts or events. Of course, most of us do not believe that people have foreknowledge, and experimentation with worlds in which there is foreknowledge necessarily involves the development of artificial societies. His work clarifies whether, in worlds in which there is foreknowledge, agents can still have choices about what to do. He shows that the answer is yes, there is still the possibility of freedom of will unless the agents' foreknowledge is total. Nevertheless, the choices they have are constrained by the need to include in their predictions of the future what is foreknown to occur. He is also able to investigate whether foreknowledge is beneficial to the survival of the agents in his artificial society (Doran 1998).

Building the model

Once the model has been designed, one can turn to its construction. This involves either writing a special computer program or using one of the many packages or toolkits that have been written to help in the development of simulations. It is almost always easier to use a package than to start afresh writing one's own program. This is because many of the issues that take time when writing a program have already been dealt with in developing the package. For example, writing code to show plots and charts from scratch is a skilled and very time-consuming task, but most packages provide some kind of graphics facility for the display of output variables. At least some of the bugs in the code of packages will have been found by the developer or subsequent users (although you should never assume that all bugs have been eliminated). The disadvantage of packages is that they are, inevitably, limited in what they can offer. There is a choice of several packages for some styles of simulation, but nothing at all is available for others. In subsequent chapters, we shall describe the available programs and comment on their merits as we consider each type of simulation.

If one has to program a simulation without the aid of a package, a question then arises about the best programming language to use. There are several desirable features for a programming language for simulation:

- The language should be well structured and allow for incremental refinement. Most simulation programming is exploratory, because usually the specification of the program develops as the problem becomes better understood. It is therefore important that the programmer can cycle easily and quickly between coding, testing and modifying the code. Interpreted languages (such as Java, Visual Basic, Python or Ruby) are often better than compiled languages (C, C++ or Pascal) in this respect, but modern compilers and programming environments mean that the difference between compilation and interpretation is now much less than it used to be.
- The language should allow easy and rapid debugging, programs should be easily instrumented and there should be good graphics libraries. Simulations generate lots of data and there needs to be an easy way of processing them into manageable form. Because so much time in writing simulation programs (as with other types of program) consists of debugging, the quality of the facilities available for testing and tracking down faults is very important.
- Once the program has been written and tested, many hundreds of runs

will be needed to carry out sensitivity analyses (see below). For this reason, the final simulation program needs to run as efficiently as possible; this implies that the language needs to be compiled rather than interpreted.

- The language should preferably be familiar to the modeller and to researchers in the simulation community, so that it is possible for others to replicate simulation results and to take over and adapt the program to their needs. It is also useful if the language is easily portable between different types of computer.

Unfortunately, these various desirable features are contradictory in their implications for the choice of a programming language for simulation. It is difficult to find one that is easy to debug, has a good graphics library, can be compiled efficiently and is portable across different computers. In practice, this means that many different languages are used for simulation, depending on the particular balance of attributes which modellers think is important for their research. However, Java, C, C++, Objective C, Prolog, Smalltalk and Lisp are probably the most common.

Verification and validation

Once one has a 'working' simulation, the next step is to check that the simulation is actually doing what one expects (Balci 1994). With a complicated computer program, it is all too easy to make errors and find that the output is the result of a mistake, rather than a surprising consequence of the model. The process of checking that a program does what it was planned to do is known as verification. In the case of simulation, the difficulties of verification are compounded by the fact that many simulations include random number generators, which means that every run is different and that it is only the distribution of results which can be anticipated by the theory. It is therefore essential to 'debug' the simulation carefully, preferably using a set of test cases, perhaps of extreme situations where the outcomes are easily predictable.

It is often useful to set up a suite of such test cases and rerun the simulation against them each time a major change is made, to check that further errors have not been introduced. To make this easier, it is also desirable to have a system that will automatically run the test suite and record the outputs, perhaps even highlighting differences between the previous run and this one, since it is these which will need attention. In order to keep a record of which version of the simulation program gave which results, a version control system, such as provided in some programming environments, can also be very useful. Chapter 9 considers these issues in more detail.

While verification concerns whether the program is working as the researcher expects it to, validation concerns whether the simulation is a good model of the target[1]. A model which can be relied on to reflect the behaviour of the target is 'valid'. Validity can be ascertained by comparing the output of the simulation with data collected from the target (see Figure 2.2). However, there are several caveats that must be borne in mind.

First, both the model and the target processes are likely to be stochastic (that is, based partly on random factors). Exact correspondence would therefore not be expected on every occasion. Whether the difference between simulation and data from the target is so large as to cast doubt on the model depends partly on the expected statistical distribution of the output measures. Unfortunately, with simulations, these distributions are rarely known and not easy to estimate.

Second, many simulations are path-dependent: the outcomes depend on the precise initial conditions chosen because these affect the 'history' of the simulation. In other words, the outcomes may be very sensitive to the precise values of some of the assumptions in the model.

Third, even if the results obtained from the simulation match those from the target, there may be some aspects of the target that the model cannot reproduce. An example is found in the world systems models considered in Chapter 3, where predictions about the growth of the world's population for the next 50 years looked plausible, but 'retrodiction' of the population to the situation 20 years in the past, using the same model and the same parameters, was completely wrong when compared with the actual world population then.

Fourth, one must not forget the possibility that the model is correct, but the data about the target are incorrect, or, more often, are themselves a result of making assumptions and estimates. For example, in Chapter 8 we shall discuss a model that aims to contribute to understanding the rise in social complexity in France 20,000 years ago. The only data against which this model can be validated are archaeological traces, which have to be subjected to a great deal of interpretation before they can be used for validation.

Another kind of difficulty arises when the model is intentionally highly abstract. It may be hard to relate the conclusions drawn from the model to any particular data from the target. For example, in Chapter 7 we shall

[1] A similar distinction is made in the philosophy of science, between internal validity (corresponding to verification) and external validity

encounter a model first proposed by Schelling (1971), which aims to explain one of the processes that could generate ethnic residential segregation. However, it is a highly abstract model and it is not clear what data could be used to validate it directly. The same issue arises with models of artificial societies, where the target is either intentionally remote from the simulation, or does not exist at all. For these models, questions of validity and of verification are hard to distinguish.

Once one has a model that appears to be valid, at least for the particular initial conditions and parameter values for which the simulation has been run, the researcher is likely to want to consider a sensitivity analysis. This aims to answer questions about the extent to which the behaviour of the simulation is sensitive to the assumptions which have been made. For example, for a model of the tax and benefit system, one might be interested in whether a small change in welfare benefit rates results in a small or a large change in the total benefits paid out by the government. It might be that if the rate of benefit is decreased, other poverty support arrangement cut in, so that the net effect on government expenditure is much smaller than the benefit decrease might suggest. Another issue that sensitivity analysis is used to investigate is the robustness of the model. If the behaviour is very sensitive to small differences in the value of one or more parameters we might be concerned about whether the particular values used in the simulation are correct.

The principle behind sensitivity analysis is to vary the initial conditions and parameters of the model by a small amount and rerun the simulation, observing differences in the outcomes. This is done repeatedly, while systematically changing the parameters. Unfortunately, even with a small number of parameters, the number of combinations of parameter values quickly becomes very large, and because each combination requires the simulation to be run again, the resources required to perform a thorough analysis can become excessive. In practice, the modeller is likely to have a good intuition about which parameters are likely to be the most important to examine.

One of the ways in which the modeller can obtain an understanding of the sensitivity of a simulation to the values of its parameters is to vary them at random, thus generating a distribution of outcomes. One or more parameters are set to values drawn from a uniform random distribution. Plotting the values of the outputs generated from many runs of the simulation will give an indication of the functional form of the relationship between the parameters and the outputs and will indicate whether small parameter changes give rise to large output variations. In effect, one is sampling the parameter space in order to build up a picture of the behaviour of the model over many different conditions.

Randomization of parameters in order to obtain a sample of conditions is one of several uses of random numbers in simulation.[2] Random numbers also have the following uses:

- They allow for all the external and environmental processes that are not being modelled (the exogenous factors) such as the effects of the job market in a simulation of household income over time. Here, the random value is substituting for an unmeasured (and perhaps unmeasurable) parameter and is equivalent to the modeller making a guess in the absence of more accurate information.
- For a similar reason, they are sometimes used to model the effects of agents' personal attributes, such as their preferences and their emotions.
- Some simulation techniques (for example, some kinds of cellular automata and agent-based models; see Chapters 7 and 8) yield different results depending on the order in which the actions of agents in the model are simulated. It is good practice to randomize the order to avoid such unwanted effects.

Whatever the reason for introducing randomness, the simulation will have to be run many times in order to observe its behaviour in a variety of conditions. Results from the simulation will need to be presented as distributions, or as means with confidence intervals. Once one has included a random element, the simulation needs to be analyzed using the same statistical methods as have been developed for experimental research (for a primer, see Box *et al*. 1978): analysis of variance to assess qualitative changes (for example, whether clusters have or have not formed) and regression to assess quantitative changes.

Publication

The final stage in simulation research is to publish the results, adding them to the stock of scientific knowledge. However, there are some particular difficulties in writing about simulation (Axelrod 1997a). Ideally, the reader

[2]Strictly speaking, computers provide only 'pseudo-random' numbers, rather than truly random numbers, but if a good generator is used there should not be any significant difference. Most simulations use large numbers of 'random' numbers and depend greatly on the accuracy of their distribution, so it is worth checking that the programming system being used for the simulation does have a good pseudo-random number generator (see Appendix C for more on this).

should be able to grasp the social science aspects of the research without being drowned in detail, but should also be able to replicate the simulation, if he or she wants to understand precisely how it works. These objectives are in tension with one another. Often, there is not space within the length of a conventional journal article or of a chapter in a book to describe a simulation sufficiently to enable replication to be carried out. One solution is to publish the code itself on the Internet. A more radical solution is to publish in one of the increasing number of electronic journals that, because they are not constrained by the costs of paper and printing, can include not only an article of standard length, but also the code, sample runs and other materials. An electronic journal also has no difficulty in publishing colour graphics, animations and other multimedia formats, which would be impossible or prohibitively expensive to reproduce on paper.[3]

Conclusion

There is still much to learn about the most effective methods for conducting simulation-based research. However, experience is growing and the lessons that have been learned can be summarized as follows:

- If the goal is understanding, use simulation to develop theories, not accurate models. Even complicated models are unlikely to reproduce the behaviour of the social world particularly well, are difficult to construct and the complexity can get in the way of discovering new relationships and principles.
- In the past, social scientists have tended to espouse either deduction (loosely, testing of sets of assumptions and their consequences) or induction (the development of theories by generalization of observations). Simulation provides a third possibility, in which one starts with a set of assumptions, but then uses an experimental method to generate data which can be analyzed inductively (Axelrod 1997a: 24). Keep in mind the need to iterate between a deductive and inductive strategy as one develops the model.
- Since many models incorporate random elements, the results of just one run cannot be relied on. It is necessary to establish that the results are robust with respect to different random values. In addition, for

[3] For an example of an electronic journal, see the *Journal of Artificial Societies and Social Simulation* at `http://jasss.soc.surrey.ac.uk/`. For an example of a multimedia report of a simulation, see the CD-ROM which accompanies the book by Epstein and Axtell (1996).

many simulations, it is important to conduct a sensitivity analysis of the effect of variations in the assumptions on which the model is based.
- While many models have the objective of simulating a specific target in the social world, it is also possible to develop models of artificial societies which may be used to investigate not just our present society, but also possible social worlds. This can be one way of developing social theories which apply generally to interacting agents.

In the following chapters, we shall be developing these methodological recommendations with respect to a number of approaches to simulation. These range from the use of 'world models' and microsimulation, which emphasize the value of simulation for prediction (for example, the effects of population growth and environmental degradation on the human world as a whole) to approaches based on multi-agent models, which emphasize the value of exploring artificial societies.