Copyrighted material – 978–1–137–32802–1

© Stewart Robinson 2014

All rights reserved. No reproduction, copy or transmission of this publication may be made without written permission.

No portion of this publication may be reproduced, copied or transmitted save with written permission or in accordance with the provisions of the Copyright, Designs and Patents Act 1988, or under the terms of any licence permitting limited copying issued by the Copyright Licensing Agency, Saffron House, 6–10 Kirby Street, London EC1N 8TS.

Any person who does any unauthorized act in relation to this publication may be liable to criminal prosecution and civil claims for damages.

The author has asserted his right to be identified as the author of this work in accordance with the Copyright, Designs and Patents Act 1988.

First published 2014 by
PALGRAVE MACMILLAN

Palgrave Macmillan in the UK is an imprint of Macmillan Publishers Limited, registered in England, company number 785998, of Houndmills, Basingstoke, Hampshire RG21 6XS.

Palgrave Macmillan in the US is a division of St Martin's Press LLC, 175 Fifth Avenue, New York, NY 10010.

Palgrave Macmillan is the global academic imprint of the above companies and has companies and representatives throughout the world.

Palgrave® and Macmillan® are registered trademarks in the United States, the United Kingdom, Europe and other countries

ISBN 978–1–137–32802–1

This book is printed on paper suitable for recycling and made from fully managed and sustained forest sources. Logging, pulping and manufacturing processes are expected to conform to the environmental regulations of the country of origin.

A catalogue record for this book is available from the British Library.

A catalog record for this book is available from the Library of Congress.

Typeset by MPS Limited, Chennai, India.

Copyrighted material – 978–1–137–32802–1

Copyrighted material – 978–1–137–32802–1

Contents

Copyrighted material – 978–1–137–32802–1

Copyrighted material – 978–1–137–32802–1

Copyrighted material – 978–1–137–32802–1

Copyrighted material – 978–1–137–32802–1

Copyrighted material – 978–1–137–32802–1

Copyrighted material – 978–1–137–32802–1

Copyrighted material – 978–1–137–32802–1

Copyrighted material – 978–1–137–32802–1

Copyrighted material – 978–1–137–32802–1

Copyrighted material – 978–1–137–32802–1

Copyrighted material – 978–1–137–32802–1

Copyrighted material – 978–1–137–32802–1

Copyrighted material – 978–1–137–32802–1

Copyrighted material – 978–1–137–32802–1

Defining Simulation: What, Why and When?

1

Key concepts

- What is simulation? *Experimentation with a simplified imitation (on a computer) of an operations system as it progresses through time, for the purpose of better understanding and/or improving that system*
- Key simulation methods used by organisations: discrete-event simulation, Monte Carlo simulation, system dynamics, agent based simulation
- Why simulate? Variability, interconnectedness and complexity of operations systems
- Advantages of simulation:
  - Over experimentation with the real system: cost, time, control of the experimental conditions, can be used when the real system does not exist
  - Over other modelling approaches: models variability, does not involve restrictive assumptions, transparent
  - The management perspective: fosters creativity, creates knowledge and understanding, visualisation and communication, consensus building
- Disadvantages of simulation: expensive, time consuming, data hungry, requires expertise, over confidence
- When to simulate? For modelling queuing systems

1.1 Introduction

The management of an airport are planning the facilities that are required in a new terminal building. Important decisions need to be made about, among other things, the number of check-in desks devoted to each airline, the size of the baggage handling system, the amount of security check positions and the number of departure gates. On top of this, the number of staff to employ and the shifts they should work need to be determined. The total investment is in the tens of millions and it is critical that these decisions are made correctly. How can the management determine the number of resources that are required in each area of the airport?

Copyrighted material – 978–1–137–32802–1

Copyrighted material – 978–1–137–32802–1

One approach would be to build the terminal and hope that it works! This seems very risky with so much at stake. Only slightly better would be to rely upon gut feel, no doubt based on some past experience with designing and managing airport terminals. A few paper calculations, or even a spreadsheet, may help, but these are unlikely to be able to handle the full complexity of the situation.

A much more effective approach is likely to be a simulation of the proposed airport terminal. This could imitate the flow of passengers and their bags through each of the key stages from arrival to departure and would act as a basis for planning airport facilities. Indeed, this is exactly what British Airways did when planning Terminal 5 at London's Heathrow airport (Beck, 2011).

Simulation models are used by many organisations to plan future facilities and to improve current ones. Manufacturing companies simulate their production lines, financial services organisations simulate their call centres, hospitals simulate the flow of patients (e.g. through emergency departments), pharmaceutical companies simulate their supply chains and transport companies simulate their delivery networks. There are many examples of simulation being used in practice.

This chapter aims to answer three questions concerning simulation:

- What exactly is a simulation?
- Why would an organisation choose to develop and use a simulation model?
- When is simulation appropriate?

1.2 What is Simulation?

This question is answered in two ways: first by defining 'simulation' and then by outlining four key simulation methods that are used by organisations.

1.2.1 Defining Simulation

Simulation models are in everyday use and so simulation is a concept that is not alien to us. For instance, weather forecasters daily show us simulations of the weather system, where we see the movement of weather fronts over the days ahead. Many of us have game consoles that simulate a whole variety of activities, enabling us to test our skills as racing drivers, adventurers and sports people. Simulations need not be computer based. Model railways and remote control boats are familiar examples of physical simulations.

So what does the term simulation mean? In its most general sense a simulation can be defined as:

An imitation of a system.

Imitation implies mimicking or copying something else. For instance, a forger imitates the work of a great artist. The Strip in Las Vegas is full of imitations: the Eiffel Tower, the New York skyline, Venice and so on. In soccer, if a

Copyrighted material – 978–1–137–32802–1

Copyrighted material – 978–1–137–32802–1

player imitates being the recipient of foul play, it is referred to as 'simulation'. Computer aided design (CAD) systems provide imitations of production facility designs and a business process map is an imitation of a business organisation. All of these can be described as a simulation in its most general sense.

There is, however, a key difference between these imitations and those examples described in the first paragraph of this section. The earlier examples involve the passage of time, whether it is the movement of trains on a track or clouds in a weather system. The second set of examples does not involve the passage of time. Hence there is a difference between the concepts of a *static simulation*, which imitates a system at a point in time, and a *dynamic simulation*, which imitates a system as it progresses through time (Law, 2007). The term simulation is mostly used in the context of dynamic simulation.

This book is concerned only with dynamic simulations. Further to this, the focus is on computer based simulations rather than physical simulations, although many of the principles that are described would still apply to the latter. Building on the previous definition, computer based dynamic simulation can be defined as follows:

An imitation (on a computer) of a system as it progresses through time.

Some aspects of this definition need exploring a little further. First, the concept of a *system* needs to be explained. In general terms a system is a collection of parts organised for some purpose (Coyle, 1996). The weather system, for instance, is a collection of parts including the sun, atmosphere, land and water, that is designed (assuming you believe in a Creator) for the purpose of maintaining life.

Checkland (1981) identifies four main classes of system:

- *Natural systems:* systems whose origins lie in the origins of the universe e.g. the atom, the Earth's weather system and galactic systems
- *Designed physical systems:* physical systems that are a result of human design e.g. a house, a car and a production facility
- *Designed abstract systems:* abstract systems that are a result of human design e.g. mathematics and literature
- *Human activity systems:* systems of human activity that are consciously, or unconsciously, ordered e.g. a family, a city and political systems

All such systems can be, and indeed are, simulated. This book, however, is concerned with simulation as it is used for modelling in private and public sector organisations. When describing and understanding these organisations two classes of system are of prime concern, that is, designed physical and human activity systems. For instance, a simulation might be developed of an automated production facility or warehouse (a designed physical system), or at the other extreme a model of regional health care delivery (a human activity

Copyrighted material – 978–1–137–32802–1

Copyrighted material – 978–1–137–32802–1

system). Many situations cannot be defined simply as either a designed physical system or a human activity system, but they lie at the interface between the two. A bank, for instance, consists of a designed physical system (the service counters, automatic tellers etc.), but it is also a human activity system where the staff and customers interact between and with one another. Indeed, many of the situations in which simulation is used lie at the interface between designed physical systems and human activity systems. For instance, service operations (banks, call centres and supermarkets), manufacturing plants, supply chains, transport systems, hospital emergency departments and military operations all involve elements of both class of system. In general terms, these systems can be referred to as *operations systems* or *operating systems*. 'An operating system is a configuration of resources [parts] combined for the provision of goods or services [purpose]' (Wild, 2002). Wild identifies four specific functions of operating systems: manufacture, transport, supply and service.

There are, of course, cases where other types of system need to be modelled as well. For instance, in a simulation of a port it may be necessary to model the tidal and weather conditions, since adverse conditions may prevent a ship from entering the port. As such, it is necessary to model, at least simply, some natural systems. In general this would involve modelling the outcome of the natural system (e.g. high winds) rather than the system itself.

A second aspect of the definition that needs exploring further is to consider the purpose of simulation models. Pidd (2009), in a more general discussion about models in management science, identifies the purpose of models as understanding, changing, managing and controlling reality. Following this theme, the purpose of a simulation can be described as obtaining a better understanding of and/or identifying improvements to a system. Improved understanding of a system, as well as the identification of improvements, is important since it informs future decision-making in the real system.

Another feature of Pidd's description of models is his emphasis on simplification. It is unlikely that a simulation of an operations system, particularly the elements of human activity, could represent its full detail. Indeed, even if it were possible, it is probably not desirable, since the time required to collect data on and model every aspect of a system would be excessive. Note that even the people of Las Vegas only built a half size replica of the Eiffel Tower, and this for a 'simple' physical structure!

A final aspect to consider is the nature of simulation model use. Some modelling approaches attempt to provide optimum answers (e.g. linear programming) or near optimum answers (e.g. heuristic methods). This is not the case for a simulation model. A simulation simply predicts the performance of an operations system under a specific set of inputs. For instance, it might predict the average waiting time for telephone customers at a call centre when a specific number of operators are employed. It is the job of the person using the simulation model to vary the inputs (the number of operators) and to run the model in order to determine the effect. As such, simulation is an experimental approach to modelling, that is, a 'what-if' analysis tool. The model

Copyrighted material – 978–1–137–32802–1

Copyrighted material – 978–1–137–32802–1

user enters a scenario and the model predicts the outcome. The model user continues to explore alternative scenarios until he/she has obtained sufficient understanding or identified how to improve the real system. As a consequence, simulation should be seen as a form of decision support system, that is, it supports decision-making rather than making decisions on behalf of the user. It should be noted, however, that most modern simulation software provide facilities for automating the experimentation process with the aim of finding an optimum scenario. These facilities and their application are discussed in Section 10.5.4.

These four aspects (operations systems, purpose, simplification and experimentation) are now added to the previous definition so that simulation is defined as:

Experimentation with a simplified imitation (on a computer) of an operations system as it progresses through time, for the purpose of better understanding and/or improving that system.

This is the nature of the simulations that are described in this book. Note that some specifics of this definition are discussed in detail in later chapters, in particular, the methods of imitating a system as it progresses through time (Section 2.2), approaches to simplification (Section 6.3) and experimentation (Chapters 9 and 10). From here on the term simulation shall be taken to mean simulation as defined above, unless otherwise stated.

1.2.2 Simulation Methods

A range of simulation methods are used by organisations. The primary approaches are discrete-event simulation, Monte Carlo simulation, system dynamics and agent based simulation. The focus of this book is on discrete-event simulation which is specifically used for modelling an organisation's operations systems. Discrete-event simulation is described in detail in Chapter 2. For the purposes of having a basic understanding of these four simulation methods, their nature and main uses are briefly described here.

Discrete-Event Simulation

Discrete-event simulation is used for modelling queuing systems. A system is represented as entities flowing from one activity (effectively a time delay) to another. Activities are separated by queues. The queues result when entities arrive at a faster rate than they can be processed by the next activity. On the surface there may seem to be a limited set of circumstances that can be described as queuing systems, but the applications are many and various. Indeed, many systems can be conceived as queuing systems, whether it is people, physical items or information that are represented by the entities moving through the system. As a result, discrete-event simulation is widely used across a whole range of organisation (Section 1.4).

Copyrighted material – 978–1–137–32802–1

Copyrighted material – 978–1–137–32802–1

Monte Carlo Simulation

Monte Carlo simulation takes its name from the famous casino in the Principality of Monaco. As the name suggests, the aim of Monte Carlo simulation is to model risk in an environment where the outcome is subject to chance. The world is conceived as a set of distributions representing variables that describe the sources of chance. The distributions are combined in some way to determine the outcome. Figure 1.1 illustrates the idea, showing three sources of chance (*a*, *b* and *c*), which are combined in the simulation (using function *f*) to produce the outcome *d*, which is the distribution of possible outcomes.

As a simple example, *a*, *b* and *c* could each represent the role of a die, giving a number between 1 and 6 with equal probability. The outcome of interest, *d*, is the total of the three roles of the die (making $f = a + b + c$). If this system is simulated once, then a single outcome is obtained. For instance, $a = 3$, $b = 2$, $c = 5$ gives $d = 10$. But this is just one of many possible outcomes. In order to determine the range of outcomes and their probability, this system could be simulated many many times to determine the distribution of outcomes for *d*. The risk of obtaining, say, a value of $d > 15$ could then be determined.

This is, of course, a very simple example for which the distribution of *d* could be determined by writing down all the combinations of die throws. The Monte Carlo approach is used widely in much more complex and meaningful environments, especially for portfolio management. In financial services, the Monte Carlo method is used to model the future of investment portfolios. In this case, the inputs are the stocks in the portfolio, each with its own distribution of possible outcomes in terms of share price in the future. The outcome is the total value of the portfolio at a point in the future. Monte Carlo simulation is also used in the pharmaceutical industry to predict the future financial performance of a set of investments in the research and development of new drugs. Strictly speaking, these simulations do not have to model the progression of time, as per our definition of simulation above, but Monte Carlo simulation is often used to simulate an outcome at some future point and even over a series of time periods (weeks, months or years).

A good introduction to the subject of Monte Carlo simulation can be found in Winston and Albright (2011).

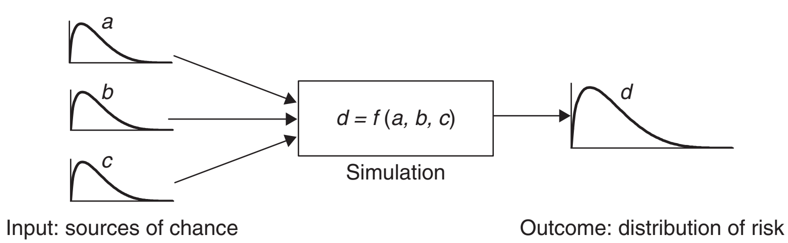

Figure 1.1 Monte Carlo simulation

Copyrighted material – 978–1–137–32802–1

Copyrighted material – 978–1–137–32802–1

System Dynamics

System dynamics is a continuous simulation approach that represents the world as a set of stocks and flows (Forrester, 1961; Coyle, 1996; Sterman, 2000). Stocks are accumulations (e.g. of items, people, money) and flows adjust the level of a stock with inflows increasing the stock and outflows reducing it. Stocks change continuously in response to the balance of the inflows and outflows from the stock; hence the need to model time continuously.

Figure 1.2 shows a simple system dynamics population model. The population, which is a stock, is increased by the birth rate, an inflow, and reduced by the death rate, an outflow. The circles represent variables and the arrows their relationships. The birth rate is determined by the size of the population and the fertility of the population. The death rate is also determined by the population and the life expectancy. Equations are used to describe the nature of these relationships.

System dynamics particularly focuses on modelling information feedback in a system (Section 1.3.1). In the example in Figure 1.2, feedback occurs between the birth rate and the population as shown by the flow into the population and the information arrow back from the population. As the birth rate increases, the population grows at a faster rate (all other things being equal) and so the birth rate increases further.

There is a very broad range of applications for system dynamics. It is particularly suited to investigating strategic issues (Morecroft, 2007). Sterman (2000) gives examples that include modelling the growth of high-tech firms, forecasting energy consumption and commodity prices, modelling supply chains and analysing business cycles.

There are a number of situations where system dynamics could be use in place of a discrete-event simulation, or *vice versa*. For instance, both are used to model supply chains (Tako and Robinson, 2012) and health care issues (Evenden et al, 2005; Rauner et al., 2005). Some recent studies have investigated the similarities and differences between discrete-event simulation and system dynamics (Morecroft and Robinson, 2005; Tako and Robinson, 2009; Tako and Robinson, 2010). In general, discrete-event simulation is more appropriate when a system needs to be modelled in detail, particularly when individual items need to be tracked through the system.

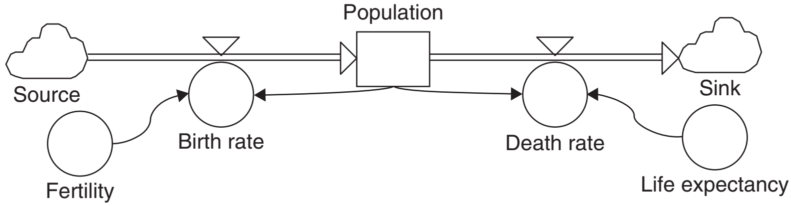

Figure 1.2 System dynamics population model

Copyrighted material – 978–1–137–32802–1

Copyrighted material – 978–1–137–32802–1

Agent Based Simulation[1]

The origins of agent based simulation lie in the desire to study complex (adaptive) systems and their emergent behaviours (Heath and Hill, 2010). The approach, that was popularised by the Sante Fe Institute through its Swarm software, has been applied across a wide range of fields for studying biological, physical and social systems, for instance. The basic idea is to model systems from the bottom-up as a set of agents, with individual behaviours, that interact over time. The aim of modelling systems in this way is to observe the behaviours, patterns and structures that emerge (Macal and North, 2010).

Macal and North (2010) describe the structure of an agent based simulation model as consisting of three elements:

- *Agents*: with attributes and behaviours
- *Agent relationships*: defining who agents interact with and how
- *Agent environment*: the environment in, and with, which the agents interact

Schelling's model of segregation is an early example of agent based simulation (in his case not on a computer) in which the dynamics of a population, that is split into two groups who aim for a desired level of segregation, is investigated (Schelling, 1971). Figure 1.3 shows an example of this model. The world is represented as a grid with light and dark grey tokens. A token desires a certain number of its neighbours to be of a similar type. If they are not, then the token moves to another space on the grid. This process continues until all tokens are satisfied with their neighbourhood. The model demonstrates that typically a much higher level of segregation is achieved than desired by each individual.

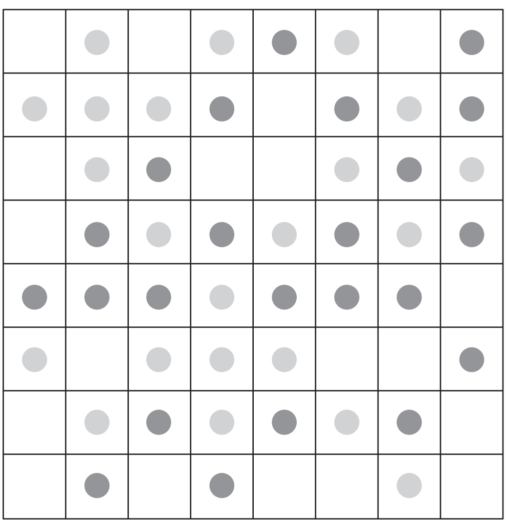

Figure 1.3 Agent based simulation: Schelling's model of segregation

Copyrighted material – 978–1–137–32802–1

Copyrighted material – 978–1–137–32802–1

Axelrod (1997) outlines a number of agent based simulation models which he uses to study phenomena in the arena of political-science. One such model investigates how cultures spread and how they remain distinct over time (Chapter 7). In a similar vein, Deffuant et al. (2002) describe a model of opinion dynamics which they use to model the influence of extremists in a population. Meanwhile, Parker and Epstein (2011) use agent based simulation for modelling epidemics. Using distributed computing their model is able to represent the outbreak of an epidemic in a population with billions of agents.

These examples of agent based simulation models are used to test the theories of the phenomena that are under investigation. In our context, we are more interested in models that are based on empirical data and that are used to aid decision-making. Recent years have seen a growing interest in the application of agent based simulation in this arena. For instance, Macal and North (2005) discuss the validation of a model (EMCAS: Electricity Market Complex Adaptive System) that was designed to aid decision-making around restructuring and deregulation of electricity power markets. Robertson and Caldart (2009) discuss the use of agent based approaches for strategic decision-making. For a detailed survey of recent examples of agent based simulation see Heath et al. (2009).

The rest of this chapter, and indeed the book, focuses on discrete-event simulation and its use for modelling operations systems. Many of the ideas, however, are more generally applicable to the other simulation methods.

1.3 Why Simulate?

In order to answer this question, three perspectives are adopted. First, the need to use simulation because of the nature of operations systems is discussed. Second, the advantages of simulation over other approaches to understanding and improving a system are described. Finally, the disadvantages of simulation are discussed, on the grounds that it is important to be cognisant of these when determining whether or not to use the approach.

1.3.1 The Nature of Operations Systems: Variability, Interconnectedness and Complexity

Many operations systems are subject to *variability*. This might be predictable variations, for instance, changing the number of operators in a call centre during the day to meet changing call volumes or planned stoppages in a production facility. It might also be variations that are unpredictable, such as, the arrival rate of patients at a hospital emergency department or the breakdown of equipment in a flexible manufacturing cell. Both forms of variability are present in most operations systems.

Operations systems are also *interconnected*. Components of the system do not work in isolation, but affect one another. A change in one part of a system leads to a change in another part of the system. For instance, if a machine is set to work faster this is likely to cause a reduction in work-in-progress up-stream and a build-up of parts down-stream.

Copyrighted material – 978–1–137–32802–1

Copyrighted material – 978–1–137–32802–1

It is often difficult to predict the effects of the interconnections in a system, especially when variability is present. Take the following example. Customers in a service process pass through three (interconnected) stages (Figure 1.4). Each stage takes exactly nine minutes. Customers arrive exactly every ten minutes. What is the average time a customer spends in the system? This is a relatively simple question to answer, since there is no variability in the system. The average time customers spend in the system is 27 minutes, in fact each customer spends exactly 27 minutes in the system.

Now assume that the times given above are averages, so customers arrive on average every 10 minutes and it takes on average 9 minutes to serve a customer at each stage. What is the average time customers spend in the system? This is not an easy question to answer since there is variability in both customer arrivals and service times: some service times will be short, others much longer; sometimes customers will arrive close together, on other occasions there will be prolonged gaps between customer arrivals. It is also expected that queues will develop between the service stages. Added to this, the range of variability around the average is not known. From extensive use of this example I have found that most people estimate that the average is still 27 minutes or maybe slightly longer. In fact, assuming a typical range of variability (a negative exponential distribution – Section 7.4.3), the average is 270 minutes.[2] The compound effect of variability and the interconnections in the system massively increase the average time customers spend in this system and make it very difficult to predict the overall performance of the system.

Many operations systems are also *complex*. It is difficult to provide an exact definition of the word complexity; an interesting discussion can be found in Gell-Mann (1994) and in relation to simulation and modelling in Brooks and Tobias (1996). For our purposes it is useful to distinguish between *combinatorial complexity* and *dynamic complexity*. Combinatorial complexity is related to the number of components in a system or the number of combinations of system components that are possible. The travelling salesman problem is a useful illustration of this. A salesperson has to make a series of visits to potential customers during a day. The aim is to find the shortest route around those customers. If there are eight 'cities' (customers) to visit, then the sales person is faced with 2,520 possible combinations of routes (this is calculated by (n–1!)/2, where n is the number of cities). As the number of cities increases, so the number of combinations grows at an increasing rate. A 16 city tour gives 6.5 x 10^{11} combinations of routes! The problem is subject to combinatorial complexity.

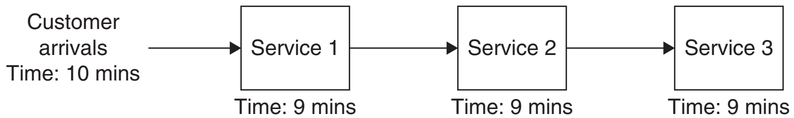

Figure 1.4 Example of an interconnected system subject to variability

Copyrighted material – 978–1–137–32802–1

Copyrighted material – 978–1–137–32802–1

Combinatorial complexity is present in some operations systems. Take, for instance, a job shop. Parts are processed through a series of machines. Once the processing is complete on one machine, a part is passed to any one of the other machines, depending on the type of part and the next process required. The more machines there are in the job shop, the more the potential interconnections. As the number of machines increases so does the interconnections at an even faster rate. Figure 1.5 shows the possible interconnections for job shops with two, three, four and five machines. There are two interconnections between any two machines since parts can move in either direction. The total number of interconnections can be calculated as $n(n–1)$, where n is the number of machines in the job shop.

On the other hand, dynamic complexity is not necessarily related to size. Dynamic complexity arises from the interaction of components in a system over time (Sterman, 2000). This can occur in systems that are small, as well as large. Systems that are highly interconnected are likely to display dynamic complexity.

Senge (1990) illustrates dynamic complexity in a supply chain with the 'beer distribution game'. This represents a simple supply chain consisting of a retailer, wholesaler and factory. The retailer orders cases of beer from the

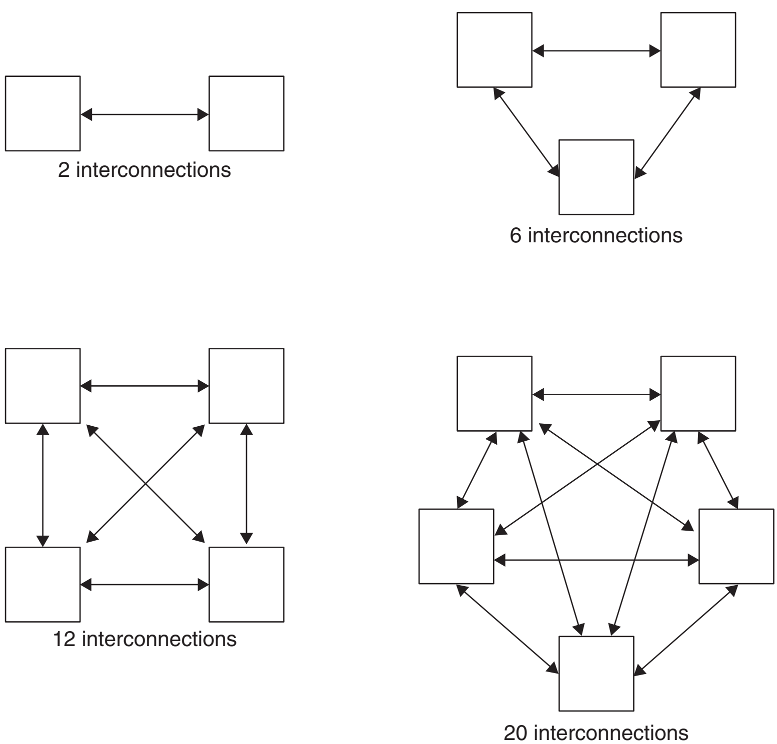

Figure 1.5 Job shop systems: interconnections and combinatorial complexity

Copyrighted material – 978–1–137–32802–1

Copyrighted material – 978–1–137–32802–1

wholesaler, who in turn orders beer from the factory. There is a delay between placing an order and receiving the cases of beer. The game demonstrates that a small perturbation in the number of beer cases sold by the retailer can cause large shifts in the quantity of cases stored and produced by the wholesaler and factory respectively. Such a system is subject to dynamic complexity.

Senge (1990) describes three effects of dynamic complexity:

- An action has dramatically different effects in the short and long run.
- An action has a very different set of consequences in one part of the system to another.
- An action leads to non-obvious consequences (counter intuitive behaviour).

These effects make it very difficult to predict the performance of a system when actions are taken, or changes are made.

The effects described above often arise because of feedback within a system. Feedback occurs when the components of a system are interconnected in a loop structure. As a result an action taken at one point in a system eventually leads to a feedback effect on that same point in the system. A simple example is a kanban system (Figure 1.6). Machine M1 feeds buffer B1. The rate at which M1 works depends upon the number of parts in B1. The smaller the inventory in B1 the faster M1 works, and *vice versa*. M1 is connected to B1 through the flow of parts and B1 is connected to M1 through the flow of information about the quantity of parts in the buffer. There is, therefore, a loop structure and so feedback occurs. For instance, if M1 is made to work faster, this increases the number of parts in B1, which in turn reduces the speed at which M1 works. If there is a delay between the request for M1 to produce at a faster rate and M1 actually delivering more parts to B1 (say, because of the need to set-up the machine to run at a faster speed), then the inventory in B1 might run out. Similarly, B1 might overfill because of a delay in M1 slowing down when requested to do so. Such undershoots and overshoots in the inventory in B1 demonstrate the dynamic complexity that can occur in this simplest of systems.

The interconnections in operations systems are often not unidirectional, and so loop structures and feedback are quite common. In particular, physical items and information often flow in opposite directions. In some cases the loop structures are very complex, involving many system components.

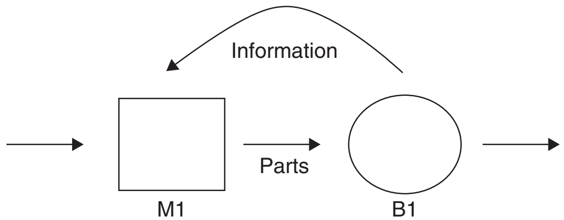

Figure 1.6 Simple kanban system demonstrating feedback

Copyrighted material – 978–1–137–32802–1

Copyrighted material – 978–1–137–32802–1

The Need for Simulation

Many operations systems are interconnected and subject to both variability and complexity (combinatorial and dynamic). Because it is difficult to predict the performance of systems that are subject to any one of variability, interconnectedness and complexity, it is very difficult, if not impossible, to predict the performance of operations systems that are potentially subject to all three. Simulation models, however, are able to explicitly represent the variability, interconnectedness and complexity of a system. As a result, it is possible with a simulation to predict system performance, to compare alternative system designs and to determine the effect of alternative designs and policies on system performance.

The methods for modelling variability in simulation are discussed in Sections 2.3 and 7.4. Section 6.2.4 discusses the need to account for the interconnections in a system. The combination of modelling variability and interconnectedness means that the complexity in a system can be represented by a simulation model.

1.3.2 The Advantages of Simulation

Simulation is not the only method of analysing and improving operations systems. In particular, it might be possible to experiment with the real system or to use another modelling approach (Pidd, 2004). What are the specific advantages of simulation over these approaches?

Simulation versus Experimentation with the Real System

Rather than develop and use a simulation model, experiments could be carried out in the real system. For instance, additional check-in desks could be placed in an airport departure area, or a change in the flow around a factory floor could be implemented. There are some obvious, and less obvious, reasons why simulation is preferable to such direct experimentation.

- *Cost.* Experimentation with the real system is likely to be costly. It is expensive to interrupt day-to-day operations in order to try out new ideas. Apart from the cost of making changes, it may be necessary to shut the system down for a period while alterations are made. Added to this, if the alterations cause the operation's performance to worsen, this may be costly in terms of loss of custom and customer dissatisfaction. With a simulation, however, changes can be made at the cost of the time it takes to alter the model and without any interruption to the operation of the real world system.
- *Time.* It is time consuming to experiment with a real system. It may take many weeks or months (possibly more) before a true reflection of the performance of the system can be obtained. Depending on the size of the model and speed of the computer, a simulation can run many times faster than real time. Consequently, results on system performance can be obtained in a matter of minutes, maybe hours. This also has the advantage

Copyrighted material – 978–1–137–32802–1

Copyrighted material – 978–1–137–32802–1

that results can be obtained over a very long time frame, maybe years of operation, if required. Faster experimentation also enables many ideas to be explored in a short time frame.

- *Control of the experimental conditions.* When comparing alternatives it is useful to control the conditions under which the experiments are performed so direct comparisons can be made. This is difficult when experimenting with the real system. For instance, it is not possible to control the arrival of patients at a hospital. It is also likely that experimentation with the real system will lead to the Hawthorne effect, where staff performance improves simply because some attention is being paid to them. In some cases the real system only occurs once, for example, a military campaign, and so there is no option to repeat an experiment. With a simulation model the conditions under which an experiment is performed can be repeated many times. The same pattern of patient arrivals can be generated time and time again, or the events that occur during a military campaign can be reproduced exactly as often as is required.
- *The real system does not exist.* A most obvious difficulty with real world experimentation is that the real system may not yet exist. Apart from building a series of alternative real world systems, which is unlikely to be practical in any but the most trivial of situations, direct experimentation is impossible in such circumstances. The only alternative is to develop a model.

Simulation versus Other Modelling Approaches

Simulations are not the only models that can be used for understanding and improving the real world. Other modelling approaches range from simple paper calculations, through spreadsheet models, to more complex mathematical programming and heuristic methods (e.g. linear programming, dynamic programming, simulated annealing and genetic algorithms). Queuing theory provides a specific class of model that looks at similar situations to those often represented by simulations, arrivals, queues and service processes (Winston, 2003). There are some reasons why simulation would be used in preference to these other methods.

- *Modelling variability.* It has already been stated that simulations are able to model variability and its effects. Meanwhile, many of the methods mentioned above are not able to do so. (It should be noted that some modelling approaches can be adapted to account for variability, but this often increases their complexity.) If the systems being modelled are subject to significant levels of variability, then simulation is often the only means for accurately predicting performance. Some systems cannot be modelled analytically. This is illustrated by Robinson and Higton (1995) who contrast the results from a 'static' analysis of alternative factory designs with a simulation. In the static analysis the variability, largely resulting from equipment failures, was accounted for by averaging their effects into the process cycle times. In the

Copyrighted material – 978–1–137–32802–1

Copyrighted material – 978–1–137–32802–1

simulation, the variability was modelled in detail. Whilst the static analysis predicted each design would reach the throughput required, the simulation showed that none of the designs were satisfactory. It is vital that variability is properly accounted for when attempting to predict performance.
- *Restrictive assumptions.* Simulation requires few, if any, assumptions, although the desire to simplify models and a shortage of data mean that some appropriate simplifications and assumptions are normally made. Many other modelling approaches require certain assumptions. Queuing theory, for instance, assumes particular distributions for arrival and service times. For many processes these distributions are not appropriate. In simulation, any distribution can be selected.
- *Transparency.* A manager faced with a set of mathematical equations or a large spreadsheet may struggle to understand, or believe, the results from the model. Simulation is appealing because it is more intuitive and an animated display of the system can be created, giving a non-expert greater understanding of, and confidence in, the model.

Of course, there are occasions when another modelling approach is appropriate and simulation is not required. Because simulation is a time consuming approach, it is recommended that it is used as a means of last resort, rather than the preferred option (Pidd, 2004). That said, simulation is often the only resort. Indeed, surveys of modelling practice demonstrate that simulation is one of the most commonly used modelling techniques (Jeffrey and Seaton, 1995; Jahangirian et al., 2010).

Simulation: The Management Perspective

Among the most compelling reasons for using simulation are the benefits that are gained by managers.

- *Fostering creativity.* 'Ideas which can produce considerable improvements are often never tried because of an employee's fear of failure' (Gogg and Mott, 1992). With a simulation, however, ideas can be tried in an environment that is free of risk. This can only help to encourage creativity in tackling problem situations.
- *Creating knowledge and understanding.* At the end of many months of simulation modelling, the manager of the organisation informed me that all of the benefits could have been obtained without the use of simulation by simply thinking about the problem in more detail. My defence lay in the fact that they would not have thought through the issues had the simulation not been there to act as a catalyst. The development and use of a simulation model forces people to think through issues that otherwise may not have been considered. The modeller seeks information, asks for data and questions assumptions, all of which lead to an improved knowledge and understanding of the system that is being simulated. Shannon (1975) recognises that the development of the model alone, without the need for

Copyrighted material – 978–1–137–32802–1

Copyrighted material – 978–1–137–32802–1

experimentation, may create sufficient understanding to bring about the necessary improvement to the real system. As the old adage states 'a problem stated is a problem half solved'.
- *Visualisation and communication.* Many good ideas have been trampled underfoot because the benefits could not be demonstrated to a senior manager. Visual simulations prove a powerful tool for communication. It may be that an idea has already been proven, but it is deemed necessary to build a simulation model in order to convince senior managers and colleagues of its validity.
- *Consensus building.* Many simulation studies are performed in the light of differing opinions as to the way forward. In the health sector, clinicians may be at odds with managers over the resources required. In a factory, managers and workers may not be in agreement over working hours and shifts. Sitting opposing parties around a simulation model of the problem situation can be a powerful means for sharing concerns and testing ideas with a view to obtaining an accommodation of views.

1.3.3 The Disadvantages of Simulation

There are a number of problems with using simulation and these must not be ignored when deciding whether or not it is appropriate.

- *Expensive.* Simulation software is not necessarily cheap and the cost of model development and use may be considerable, particularly if consultants have to be employed.
- *Time consuming.* It has already been stated that simulation is a time consuming approach. This only adds to the cost of its use and means that the benefits are not immediate.
- *Data hungry.* Most simulation models require a significant amount of data. This is not always immediately available and where it is, much analysis may be required to put it in a form suitable for the simulation.
- *Requires expertise.* Simulation modelling is more than the development of a computer programme or the use of a software package. It requires skills in, among other things, conceptual modelling, validation and statistics, as well as skills in working with people and project management. These are the skills that are discussed in this book. This expertise is not always readily available.
- *Over confidence.* There is a danger that anything produced on a computer is seen to be right. With simulation this is further exacerbated with the use of an animated display, giving an appearance of reality. When interpreting the results from a simulation, consideration must be given to the validity of the underlying model and the assumptions and simplifications that have been made.

1.4 When to Simulate

As stated in Section 1.2.2, discrete-event simulation is used for modelling queuing systems. This provides a general structure that can model a wide range of systems. As such, it is impossible to give a full list of applications for which simulation

Copyrighted material – 978–1–137–32802–1

Copyrighted material – 978–1–137–32802–1

might be used. It is, however, useful to give some indication of the range of systems that can be modelled. Banks et al. (1996) suggest the following list:

- Manufacturing systems
- Public systems: health care, military, natural resources
- Transportation systems
- Construction systems
- Restaurant and entertainment systems
- Business process reengineering/management
- Food processing
- Computer system performance

There are no doubt other applications that can be added to this list, for instance, service and retail systems.

1.5 Conclusion

This chapter discusses the nature of simulation that is described in this book. While a specific definition of simulation for modelling operations systems is provided, it is also shown that the term 'simulation' has many meanings. Four different simulation methods are outlined, whilst recognising that the focus of this book is on one of those methods, discrete-event simulation. The reasons for using simulation are discussed based on the nature of operations systems and the advantages of simulation. The latter describes why simulation is often preferable to other improvement approaches that could be adopted. The disadvantages of simulation are also identified. Finally, some common application areas for simulation modelling are listed. Having set the scene, the next chapter describes how a simulation model works by showing how the progression of time and variability are modelled.

Exercises

E1.1 Think of situations where simulation could be used, for instance, from day-to-day life, a place of study or work. What aspects of each situation make simulation appropriate?

E1.2 Take a typical operations system, preferably one that can be observed (e.g. a bank or supermarket), and identify the elements of variability, interconnectedness and complexity.

E1.3 There are many case studies describing the application of simulation to real problems. Obtain and read some simulation case studies. Why was simulation used? What benefits were obtained? Some journals that often publish simulation case studies are: *IIE Solutions*, *Interfaces*, *OR Insight* and the *Journal of the Operational Research Society*. The Winter Simulation Conference proceedings (www.wintersim.org) include many case studies. Simulation software suppliers also publish case studies on their web sites (Section 3.3.3).

Copyrighted material – 978–1–137–32802–1

Copyrighted material – 978–1–137–32802–1

E1.4 The paper by Tako and Robinson (2010, p. 787) describes a case study of modelling the UK prison population as follows:

> [T]wo types of offenders are considered, petty and serious. There are in total 76,000 prisoners in the system, of which 50,000 are petty and 26,000 serious offenders. Offenders enter the system as first time offenders and receive a sentence depending on the type of offence; on average 3,000 petty offenders enter per year versus 650 serious offenders per year. Petty offenders receive a shorter sentence (on average 5 years vs. 20 years for serious offenders). After serving time in prison they are released. A proportion of the released prisoners re-offend and go back to jail (recidivists) after 2 years (on average), whereas the rest are rehabilitated.

Discuss how the four different simulation methods (discrete-event simulation, Monte Carlo simulation, system dynamics and agent based simulation) might be used to represent this problem.

Notes

1. Parts of this section are based on Robinson, S. (2013). Conceptual Modeling for Simulation. *Proceedings of the 2013 Winter Simulation Conference* (R. Pasupathy, S.-H. Kim, A. Tolk, R. Hill, and M. E. Kuhl, eds). IEEE, Piscataway, NJ, pp. 377–388.
2. This is an exact result obtained from queuing theory formula, which can be used if it is assumed that arrivals and service times are distributed according to a negative exponential distribution. In this case the arrival rate $\lambda=1/10$ and the service rate $\mu=1/9$. For these values, standard queuing formula give the average time in the system for a single server queue as 90 minutes. This can simply be multiplied by three to give the average for the three stage queuing process. For details of the queuing formula for this problem see Winston (2003).

References

Axelrod, R. (1997). *The Complexity of Cooperation: Agent-Based Models of Competition and Collaboration*. Princeton University Press, Princeton, NJ.

Banks, J., Carson, J.S. and Nelson, B.L. (1996). *Discrete-Event System Simulation, 2nd ed*. Prentice-Hall, Upper Saddle River, NJ.

Beck, A. (2011). Modelling Passenger Flows in Heathrow Terminal 5. *Journal of Simulation* 5 (2), pp. 69–76.

Brooks, R.J. and Tobias, A.M. (1996). Choosing the Best Model: Level of Detail, Complexity and Model Performance. *Mathematical and Computer Modelling*, 24 (4), pp. 1–14.

Checkland, P. (1981). *Systems Thinking, Systems Practice*. Wiley, Chichester, UK.

Copyrighted material – 978–1–137–32802–1

Copyrighted material – 978–1–137–32802–1

Coyle, R.G. (1996). *System Dynamics Modelling: A Practical Approach*. Chapman & Hall, London.

Deffuant, G., Amblard, F., Weisbuch, G. and Faure, T. (2002) How can Extremism Prevail? A Study Based on the Relative Agreement Interaction Model. *Journal of Artificial Societies and Social Simulation*, 5 (4) 1. http://jasss.soc.surrey.ac.uk/5/4/1.html accessed April 2013.

Evenden, D., Harper, P.R., Brailsford, S.C. and Harindra, V. (2005) System Dynamics Modelling of Chlamydia Infection for Screening Intervention Planning and Cost Benefit Estimation. *IMA Journal of Management Mathematics*, 16, (3), pp. 265–279.

Forrester, J.W. (1961). *Industrial Dynamics*. MIT Press, Cambridge, MA.

Gell-Mann, M. (1994). *The Quark and the Jaguar: Adventures in the Simple and the Complex*. Abacus, London.

Gogg, T. and Mott, J. (1992). *Improve Quality and Productivity with Simulation*. JMI Consulting Group, Palos Verdes Pnsl., CA.

Heath, B., Hill, R. and Ciarallo, F. (2009). A Survey of Agent-Based Modeling Practices (January 1998 to July 2008). *Journal of Artificial Societies and Social Simulation*, 12 (4) 9. http://jasss.soc.surrey.ac.uk/12/4/9.html accessed April 2013.

Heath, B.L. and Hill, R.R. (2010). Some Insights into the Emergence of Agent-Based Modelling. *Journal of Simulation*, 4 (3), pp. 163–169.

Jahangirian, M., Eldabi, T., Naseer, A., Stergioulas, LK. and Young, T. (2010). Simulation in Manufacturing and Business: A Review. *European Journal of Operational Research*, 203 (1), pp. 1–13.

Jeffrey, P. and Seaton, R. (1995). The Use of Operational Research Tools: A Survey of Operational Research Practitioners in the UK. *Journal of the Operational Research Society*, 46 (7), pp. 797–808.

Law, A.M. (2007). *Simulation Modeling and Analysis, 4th ed*. McGraw-Hill, New York.

Macal, C.M. and North, M.J. (2005). Validation of an Agent-based Model of Deregulated Electric Power Markets. *Proceedings of North American Computational Social and Organization Science (NAACSOS)*.

Macal, C.M. and North, M.J. (2010). Tutorial on Agent-Based Modelling and Simulation. *Journal of Simulation*, 4 (3), pp. 151–162.

Morecroft, J. (2007). *Strategic Modelling and Business Dynamics: A Feedback Systems Approach*. Wiley, Chichester, UK.

Morecroft, J. and Robinson, S. (2005). Explaining Puzzling Dynamics: Comparing the Use of System Dynamics and Discrete-Event Simulation. *The 23rd International Conference of the System Dynamics Society*. www.systemdynamics.org/conferences/2005/proceed/papers/MOREC107.pdf accessed 4 April 2014.

Parker, J. and Epstein, J.M. (2011). A Distributed Platform for Global-Scale Agent-Based Models of Disease Transmission. *ACM Transactions on Modeling and Computer Simulation*, 22 (1), pp. 2:1–2:25.

Copyrighted material – 978–1–137–32802–1

Copyrighted material – 978–1–137–32802–1

Pidd, M. (2004). *Computer Simulation in Management Science, 5th ed.* Wiley, Chichester, UK.

Pidd, M. (2009). *Tools for Thinking: Modelling in Management Science, 3rd ed.* Wiley, Chichester, UK.

Rauner, M.S., Brailsford, S.C. and Flessa, S. (2005) Use of Discrete-Event Simulation to Evaluate Strategies for the Prevention of Mother-to-Child Transmission of HIV in Developing Countries. *Journal of the Operational Research Society*, 56, (2), pp. 222–233.

Robertson, D.A. and Caldart, A.A. (2009). *The Dynamics of Strategy: Mastering Strategic Landscapes of the Firm*. Oxford University Press.

Robinson, S. and Higton, N. (1995). Computer Simulation for Quality and Reliability Engineering. *Quality and Reliability Engineering International*, 11, pp. 371–377.

Schelling, T.C. (1971). Dynamic Models of Segregation. *Journal of Mathematical Sociology*, 1, pp. 143–186.

Senge, P.M. (1990). *The Fifth Discipline: The Art and Practice of the Learning Organization*. Random House, London.

Shannon, R.E. (1975). *Systems Simulation: The Art and Science*. Prentice-Hall, Englewood Cliffs, NJ.

Sterman, J.D. (2000). *Business Dynamics: Systems Thinking and Modeling for a Complex World*. McGraw-Hill, New York.

Tako, A.A. and Robinson, S. (2009). Comparing Discrete-Event Simulation and System Dynamics: Users' Perceptions. *Journal of the Operational Research Society*, 60 (3), pp. 296–312.

Tako, A.A. and Robinson, S. (2010). Model Development in Discrete-Event Simulation and System Dynamics: An Empirical Study of Expert Modellers. *European Journal of Operational Research*, 207, pp. 784–794.

Tako, A.A. and Robinson, S. (2012). The Application of Discrete Event Simulation and System Dynamics in the Supply Chain Context. *Decision Support Systems*, 52 (4), pp. 802–815.

Wild, R. (2002). *Operations Management, 6th ed.* Continuum, London.

Winston, W.L. (2003). *Operations Research: Applications and Algorithms, 4th ed.* Duxbury Press, Belmont, CA.

Winston, W.L. and Albright, S.C. (2011). *Practical Management Science, 4th ed.* Cengage Learning, Independence, Kentuky.

Copyrighted material – 978–1–137–32802–1

Copyrighted material – 978–1–137–32802–1

Index

Copyrighted material – 978–1–137–32802–1

Copyrighted material – 978–1–137–32802–1

Copyrighted material – 978–1–137–32802–1

Copyrighted material – 978–1–137–32802–1

Copyrighted material – 978–1–137–32802–1

Copyrighted material – 978–1–137–32802–1

Copyrighted material – 978–1–137–32802–1

Copyrighted material – 978–1–137–32802–1

Copyrighted material – 978–1–137–32802–1

Copyrighted material – 978–1–137–32802–1

Copyrighted material – 978–1–137–32802–1

Copyrighted material – 978–1–137–32802–1

Copyrighted material – 978–1–137–32802–1

Copyrighted material – 978–1–137–32802–1

Copyrighted material – 978–1–137–32802–1

Copyrighted material – 978–1–137–32802–1

Copyrighted material – 978–1–137–32802–1

Copyrighted material – 978–1–137–32802–1

Copyrighted material – 978–1–137–32802–1