Abstract  System dynamics is an approach for thinking about and simulating situations and organisations of all kinds and sizes by visualising how the elements fit together, interact and change over time. This chapter, written by John Morecroft, describes modern system dynamics which retains the fundamentals developed in the 1950s by Jay W. Forrester of the MIT Sloan School of Management. It looks at feedback loops and time delays that affect system behaviour in a non-linear way, and illustrates how dynamic behaviour depends upon feedback loop structures. It also recognises improvements as part of the ongoing process of managing a situation in order to achieve goals. Significantly it recognises the importance of context, and practitioner skills. Feedback systems thinking views problems and solutions as being intertwined. The main concepts and tools: feedback structure and behaviour, causal loop diagrams, dynamics, are practically illustrated in a wide variety of contexts from a hot water shower through to a symphony orchestra and the practical application of the approach is described through several real examples of its use for strategic planning and evaluation.

2.1  Introduction

In this chapter the basic concepts of system dynamics are introduced. The approach helps you to develop a strategic overview of organisations and to visualise how the parts of a business, industry or society fit together and interact to create dynamics and change through time. You will first learn how to draw and interpret causal loop diagrams that show the main interdependencies in real-world problem situations. Examples based on transportation policy, hotel showers, drug-related crime and
orchestra management illustrate how such diagrams reveal the big picture from underlying operating detail. Next you will see the steps required to create a full-blown system dynamics simulation model. Simulation is used to investigate the performance over time that arises from interlocking feedback loops. An example from the airline industry examines the dynamics of growth strategy and illustrates a fundamental tenet of the field that ‘feedback structure gives rise to dynamical behaviour’.

2.1.1 Ways of Interpreting Situations in Business and Society

The idea there is an enduring structure to business and social systems, that somehow predetermines achievable futures, is not necessarily obvious. Some people argue that personalities, ambition, chance, circumstance, fate, and unforeseen events hold the keys to the future in an uncertain world. But an interpretation of what is happening around you depends on your perspective. What appears to be chance may, from a different perspective, have a systemic cause. For example, when driving on a busy highway you may experience sporadic stops and starts. Does that mean you are at the mercy of random events like breakdowns or accidents? Not necessarily. Trapped in a car at ground level you don’t see the waves of traffic that arise from the collective actions of individual drivers as they try to maintain a steady speed while keeping a safe distance from the car in front. There is an invisible structure to the ‘system’ of driving on a crowded motorway that causes sporadic stops and starts, without the need for accidents (though, of course, they do happen too). You can sense such structure, or at least something systemic, in the pattern of traffic density (alternating bands of congestion and free flow) observable from a nearby hillside overlooking the motorway, where you have the benefit of an overview. The same benefit of perspective applies to all kinds of business and social problems. In particular there are two contrasting perspectives that people bring to bear on policy and strategy development: an event-oriented approach and a feedback (or joined-up) approach. In many ways they are polar extremes.

2.1.2 Event-Oriented Thinking

An event-oriented perspective is pragmatic, action oriented, alluringly simple and often myopic. Figure 2.1 depicts this mindset in the abstract. It reflects a belief that problems are sporadic, stemming from uncontrollable events in the outside world. Life is capricious. Events come out of the blue or at least there is no time to worry about their causes. What’s important is to fix the problem as soon as possible.

The typical thinking style here is linear – from problem-as-event to solution-as-fix. The problem presents itself as a discrepancy between an important shared goal
and a capricious current situation. Through decision and action those responsible for the shared goal arrive at a solution and then move on to the next problem. Event oriented thinking is widespread and often compelling. It can lead to swift and decisive action. But there are limitations to this open-loop, fire-fighting mode of intervention.

Consider a few practical examples depicted in Fig. 2.2. Binge drinking is often in the news. Among other things it leads to unruly behaviour in towns and cities late at night. A local solution is to deploy more police to arrest the main troublemakers. Such an approach may reduce violence and accidents on a given night, but it does not get to grips with why people are binge drinking in the first place. Similarly a quick fix solution to drug related crime is to deploy more police in order to seize drugs and arrest drug dealers, but that does not deter addicts. In a totally different area of public policy, traffic congestion is a chronic problem for motorists and transportation planners alike. One practical solution is to build new roads, an approach that does work, at least in the short run. However experience suggests that in the long run congestion returns – as in the case of the M25 orbital motorway around Greater London, originally a six lane highway with a circumference of 160 miles, completed in the mid-1980s. More than 20 years later there are sections with 12 lanes and still it is overcrowded.
2.1.3 Feedback Systems Thinking

A feedback approach is different from event-oriented thinking because it strives for solutions that are ‘sympathetic’ with their organisational and social environment. Problems do not stem from events, and solutions are not implemented in a vacuum. Instead problems and solutions coexist and are interdependent. There is a long history to these ideas. They were lucidly brought to the attention of policymakers and business leaders in Senge’s influential book *The Fifth Discipline* published in 1990. He presents four core ‘disciplines’ of successful organisational change that include team learning, shared vision, personal mastery and mental models. The fifth discipline is systems thinking which, by uniting the other disciplines, provides concepts and tools to visualise complexity and better understand sources of resistance to organisational change.² Although there are no formal simulation models or equations in the *Fifth Discipline* its approach to systems was inspired by the field of system dynamics, beginning with Forrester’s seminal book *Industrial Dynamics* published in 1961 and further developed in the many ways described in the 50th anniversary review of the field (Sterman 2007). Moreover, system dynamics itself can be meaningfully placed among intellectual traditions that contributed to the evolution of feedback concepts in the social sciences, as described by Richardson (1991).

2.1.4 An Illustration of Feedback Systems Thinking

Consider Fig. 2.3, which is a causal loop diagram of factors contributing to road use and traffic congestion (Sterman 2000). The rules for constructing such a diagram are introduced later, but for now just focus on the cause and effect links that depict far-reaching interdependencies between highway capacity and traffic volume. Four feedback loops are shown. The top loop depicts road construction by the government agency responsible for transportation. As motorists experience an increase in travel time relative to desired travel time (the amount of time they are willing to spend on travel) there is growing pressure on planners to reduce congestion. This pressure leads to road construction which, after a time delay of several years, results in more highway capacity. More highway capacity reduces travel time as motorists are able to reach their destinations more quickly on less crowded roads. The four links described so far make a closed feedback loop labelled capacity expansion. Interestingly this loop includes an event-oriented link from ‘pressure to reduce congestion’ to road construction which is similar to the connection in Fig. 2.2 from congestion to ‘build new roads’. But this isolated connection is now placed in context of many other factors, or side-effects, deemed relevant to the big picture.

---

² Here I use the term ‘feedback systems thinking’ to avoid confusion with ‘systems thinking’ often used in connection with Soft Systems Methodology (SSM), as described elsewhere in this book.
One important side effect is shown in the middle loop labelled ‘discretionary trips’. Here a reduction in travel time leads to an increase in the attractiveness of driving. Attractiveness itself depends on a variety of factors including desired travel time, adequacy of public transit and public transit fare. The greater the attractiveness of driving then (eventually) the more trips per day taken by motorists, the more traffic volume and the higher the travel time, thereby closing the loop. Here already is a vital side effect that can, in the medium to long term, defeat the objective of new road building programmes aimed at reducing congestion. Bigger and better roads make it more attractive to drive. So people make extra journeys. This particular side effect is largely responsible for the failure of London’s M25 orbital motorway to relieve traffic congestion in and around the Greater London area, as drivers took to commuting regularly between places they would otherwise seldom visit.

The lower middle loop shows a related side effect labelled ‘extra miles’. For the same attractiveness reasons drivers not only make extra journeys, they also take much longer journeys. The aggregate traffic effect is similar. Traffic volume increases leading to longer journey times.

The bottom loop labelled ‘take the bus’ shows another side effect, a potential long-term impact from public transit. Here, as the attractiveness of driving increases, public transit ridership decreases, causing cars per person to increase. (The direction of these causal effects can be read accurately from the diagram, but first you have to be familiar with the meaning of the ‘+’ and ‘−’ signs near the arrow.
heads which is explained later in the chapter.) With more cars per person there are more cars in the region and traffic volume increases, thereby closing the bottom loop.

If you reflect for a moment on the picture as a whole you realise it is a sophisticated view of the congestion problem. There are 15 concepts connected by 19 links. A lot of complexity is condensed into a small space. Compare the picture with the single stark arrow in Fig. 2.2 from an event oriented perspective. Obviously there is much more to think about and discuss in the causal loop diagram. Such richness is typical of good feedback systems thinking. The approach gives pause for thought by showing that often there is more going on (in public policy or in business strategy) than people first recognise. Where exactly to draw the boundary on the factors to include is a matter of judgement and experience. Usually there is no one right answer and therefore the process of constructing diagrams, and tying them to a dynamic phenomenon, is important too.

### 2.1.5 A Shift of Mind

People responsible for strategy development and facing problem situations often have in mind partial and conflicting views of these situations. It is therefore well worth spending time to capture their individual perspectives, develop an overview, share the big picture and thereby try to anticipate the ramifications, knock-on consequences, and side-effects of strategic change. These are the advantages of feedback systems thinking. In *The Fifth Discipline*, Peter Senge (1990) makes the point that feedback systems thinking is a ‘shift of mind’, a new way of interpreting the business and social world, and a kind of antidote to silo mentalities and narrow functional perspectives often fostered (inadvertently) by organisations and by our tendency to carve-up problems for analysis. Figure 2.4 summarises this shift of mind. Essentially problems and solutions are viewed as intertwined. The typical thinking style here is circular – starting from a problem, moving to a solution and then back to the problem. The important point, as shown on the right of the figure, is that problems do not just spring from nowhere, demanding a fix. They are a consequence of the cumulative effect of previous decisions and actions, sometimes intentional, but often with hidden side-effects.

As before, a problem presents itself as a discrepancy between an important goal and the current situation. Those responsible for achieving the goal arrive at a solution in the form of a decision leading to action and results that change the current situation. If all goes to plan then the current situation moves closer to the goal, the size of the discrepancy is reduced and the problem is alleviated. But this feedback response is not viewed as a once-and-for-all fix. It is part of a continual process of ‘managing’ the situation in order to achieve an agreed goal (or goals). Moreover, there is a recognition that other influences come to bear on the current situation. There are other stakeholders, with other goals, facing other
situations and taking their own corrective action as shown on the left of Fig. 2.4. The performance of the enterprise as a whole arises from the interplay of these interlocking feedback processes, just as we saw in the transport example where the stakeholders included motorists, transportation planners, bus companies and bus passengers.

### 2.2 Concepts and Tools of System Dynamics

In this section we review the main concepts and tools used in system dynamics modelling. The material is divided into six parts labelled 2.2.1 through 2.2.6. Part 2.2.1 introduces the rules for constructing causal loop diagrams and for interpreting their likely problematic behaviour through time. A familiar example of a slow-to-respond hot water shower illustrates the central idea that feedback structure determines dynamic behaviour. Part 2.2.2 says more about causal loops and dynamic behaviour with an example in drug-related crime. Basic tips are then presented for constructing causal loop diagrams that are both conceptually and visually clear. Part 2.2.3 introduces additional modelling symbols necessary to build full-blown simulation models. Part 2.2.4 presents the equations for a simple drug-related crime model and uses them to illustrate principles of equation formulation. Part 2.2.5 shows system dynamics in action with an application to growth strategy for a low cost airline. The example includes causal loop diagrams, equation formulations, simulations and a gaming simulator that readers can operate themselves. Finally, in a very different organisational setting, Part 2.2.6 presents causal loops from a study of symphony orchestras and their success.
2.2.1 Causal Loop Diagrams, Feedback Structure and Behaviour Through Time

A causal loop diagram is a visual tool for the feedback systems thinker. As in the transportation example, such diagrams show cause and effect relationships and feedback processes. All causal loop diagrams are constructed from the same basic elements: words, phrases, links and loops – with special conventions for naming variables and for depicting the polarity of links and loops. Figure 2.5 is a very simple causal loop diagram, just a single loop, connecting hunger and amount eaten. Deliberately there is very little detail. Imagine the situation for yourself. You are hungry, so you eat. How would you describe the process that regulates food intake? Common sense and experience says there is a relationship between hunger and amount eaten and this is shown by two causal links. In the top link hunger influences amount eaten, while in the bottom link amount eaten has a reverse influence on hunger. Each link is assigned a polarity, either positive or negative. A positive ‘+’ link means that if the cause increases then the effect increases too. So an increase in hunger causes an increase in the amount eaten. A negative ‘−’ link means that if the cause increases then the effect decreases. So an increase in the amount eaten causes a decrease in hunger. In fact the assignment of link polarity is just a bit more sophisticated. In general it is better to imagine the effect (whether an increase or decrease) relative to what it would otherwise have been, in the absence of an increase in the cause. This turns out to be a more robust test. In any case the two concepts, hunger and amount eaten, are mutually dependent,

![Causal Loop Diagram](image)

**Fig. 2.5** Simple causal loop diagram of food intake

---

3 This more sophisticated assignment of link polarity works equally well for normal straightforward causal links and for links that correspond to stock accumulation processes. The distinction will become clear when stock accumulation is introduced as a vital concept for modelling and simulating dynamical systems.
and this two-way dependence is shown as a closed feedback loop. The feedback loop represents, in outline, the control of food intake.

There are a few more details to explain in the diagram. The bottom link contains a box labelled ‘DELAY’. This symbol shows a time delay in a causal link where a given cause leads to an effect, but not immediately. There is a lag. So here the more you eat the less hungry you feel, but it takes a while for hunger pangs to diminish. Such time delays add dynamic complexity because cause and effect is less obvious. Where eating is concerned a time delay of 20 min or so can make it much more difficult to regulate food intake. Overeating is a common result. In the centre of the diagram there is another special symbol, a ‘B’ inside a small curved arrow, a loop identifier to indicate a balancing feedback loop. Generally speaking a feedback loop can be either balancing or reinforcing. The names give a clue about the way the feedback process operates. In a balancing loop a change in the condition of a given variable leads to a counteracting or balancing change when the effects are traced around the loop.

A simple thought experiment illustrates the idea. Imagine you take a long walk and return home feeling hungry. Hunger rises and the feedback loop swings into action. Amount eaten rises and eventually hunger declines. The feedback effect of the loop is to counteract the original rise in hunger, which is a balancing process. By comparison a reinforcing loop amplifies or reinforces change. In a realistic multi-loop system, such as the transport example mentioned earlier, behaviour through time arises from the interplay of balancing and reinforcing loops. So it is useful when interpreting a web of causal connections to identify the main loops as a way of telling a story of what might unfold. At the same time it is a good discipline to name each loop with a mnemonic for the underlying feedback process. In Fig. 2.5 the balancing loop is called ‘control of food intake’. Similarly in Fig. 2.3 a feedback view of road congestion is depicted vividly as the interplay of balancing loops for capacity expansion, discretionary trips, extra miles and ‘take the bus?’.

### 2.2.1.1 Feedback Structure and the Dynamics of a Slow-to-Respond Shower

Causal loop diagrams are a stepping-stone to interpreting and communicating dynamics or performance through time. The best way to appreciate this point is to see a worked example. Here I present a hot water shower like the one at home or in a hotel room. In this example we start from dynamics of interest and then construct a causal loop diagram that is capable of explaining the dynamics. Our analysis begins with a time chart as shown in Fig. 2.6. On the vertical axis is the water temperature at the shower head and on the horizontal axis is time in seconds. Imagine it is a hot summer’s day and you want to take a nice cool shower at 25°C. When you step into the cubicle the shower is already running but the water temperature is much too cold. The time chart shows three alternative time paths or trajectories for the water temperature labelled ‘ideal, ‘common sense’, and ‘most likely’. The ideal outcome is that you quickly adjust the tap setting by just the right amount and the water temperature immediately rises to the desired 25° after which it remains rock steady.
You are comfortably cool. Common sense says this ideal can’t happen because, like most showers, this one is slow to respond. There is a time delay of a few seconds between adjusting the tap and a change in the water temperature. To begin with the common sense trajectory is flat and the water temperature remains too cold. Then after a while the temperature begins to rise and quite soon settles at the desired 25°C. Unfortunately experience contradicts this common sense outcome. The most likely trajectory is much different. Again the temperature starts too cold. You adjust the tap and gradually the temperature rises. After a few seconds the temperature is just right. But annoyingly it continues to rise. Before long you are much too hot, so you reverse the tap. It makes no immediate difference. So you reverse the tap even more. At last the temperature begins to fall and after a few more seconds you are again comfortably cool at 25°C. However, your comfort is short-lived as the water temperature continues to fall and you are right back where you started – too cold. The cycle continues from cold to hot and back again.

The most likely trajectory is a classic example of puzzling dynamics, performance over time that is both unintended and surprising. Who would deliberately set-out to repeatedly freeze and scald themselves? The feedback systems thinker looks for the structure, the web of relationships and constraints involved in operating a shower, that causes normal people to self-inflict such discomfort. It is clear from Fig. 2.6 that the dynamic behaviour is essentially goal seeking. The shower taker wants the water temperature to be 25°C, but the actual water temperature varies around this target. The feedback structure that belongs with such fluctuating behaviour is a balancing loop with delay, and that’s exactly what we are looking for in modelling or representing the shower ‘system’. This notion of having in mind a structure that fits (or might fit) observed dynamics is common in system dynamics modelling. It is known formally as a ‘dynamic hypothesis’, a kind of preliminary guess at the sort of relationships likely to explain a given pattern of behaviour through time.

Figure 2.7 shows a causal loop diagram for a slow-to-respond shower. First consider just the words. Five phrases are enough to capture the essence of the troublesome shower: desired water temperature, actual water temperature, temperature gap, the flow of hot water and the flow of cold water. Next consider the causal links. The temperature gap depends on the difference between desired and actual water temperature.
The existence of a temperature gap influences the flow of hot water. This link represents the decision making and subsequent action of the shower taker. You can imagine a person turning a tap in order to change the flow of hot water and to get comfortable. The flow of hot water then influences the actual water temperature, but with a time delay because the shower is slow-to-respond. Also shown is a separate inflow of cold water, represented as a link on the left. The water temperature obviously depends on both water flows, hot and cold.

The end result is a balancing feedback loop, labelled ‘comfort seeking’, which is just what we are looking for to explain cyclical behaviour. The loop-type can be confirmed by adding signs (positive or negative) to each link and telling a ‘story’ about the process of temperature adjustment around the loop. For convenience imagine the desired water temperature is greater than actual at time zero – in other words the shower taker feels too cold and the temperature gap is greater than zero. Now consider the polarity of the first link. If the temperature gap increases then the flow of hot water becomes greater than it would otherwise have been. This is a positive link according to the polarity conventions. In the second link, if the flow of hot water increases, then the actual water temperature increases, albeit with a delay. This too is a positive link. (Note that in making the polarity assignment the flow of cold water, which also affects water temperature, is assumed to be held constant.) In the third and final link, if the water temperature increases then the temperature gap becomes smaller than it would otherwise have been. This is a negative link according to the polarity conventions. The overall effect around the loop is for an increase in the temperature gap to result in a counteracting decrease in the temperature gap, which is the signature of a balancing loop.

Incidentally, there is another way to work out loop polarity besides telling a story around the loop. It is also possible to simply count the number of negative
links around the loop. An odd number of negative links (1, 3, 5, …) signifies a balancing loop while an even number of links (0, 2, 4, …) signifies a reinforcing loop. The reason this rule-of-thumb works is that any story about propagation of change around a loop will result in a counteracting effect for an odd number of negative links and a reinforcing effect for an even number. In this case there is one negative link around the loop (between actual water temperature and the temperature gap) and so it is a balancing loop. The other negative link in the diagram (between flow of cold water and actual water temperature) does not count since it is not part of the closed loop.

2.2.1.2 Processes in a Shower ‘System’

A typical causal loop diagram shows a lot about connectivity in a small space. It is a purely qualitative model, a sketch of cause and effect, particularly good for highlighting feedback loops that contribute to dynamics and to dynamic complexity. Usually there are many practical operating details that lie behind the scenes of causality. Although not shown in the diagram it is important to be aware of this detail, particularly when building an algebraic simulator of the same feedback structure. Then it is vital to be clear and precise about how such links actually work in terms of underlying behavioural responses, economic and social conventions, and physical laws. It is also important to know the numerical strength of the effects. This skill of seeing the big picture while not losing sight of operating detail is a hallmark of good system dynamics practice, known as ‘seeing the forest and the trees’ (Senge 1990; Sherwood 2002). It is a skill well-worth cultivating.

One way to forge the connection from feedback loops to operations is to ask yourself about the real-world processes that lie behind the links. In the case of the shower there is an interesting mixture of physical, behavioural and psychological processes. Take for example the link from the flow of hot water to actual water temperature. What is really going on here? The diagram says the obvious minimum: if the flow of hot water increases then sooner or later, and all else remaining the same, the actual water temperature at the shower head increases too. The sooner-or-later depends on the time delay in the hot water pipe that supplies the shower, which is a factor that can be estimated or measured. But how much does the temperature rise for a given increase in water flow? The answer to that question depends on physics and thermodynamics – the process of blending hot and cold water. In a simulation model you have to specify the relationship with reasonable accuracy. You do not necessarily need to be an expert yourself, but if not then you should talk with someone who knows (from practice or theory) how to estimate the water temperature that results from given flows of hot and cold water – a plumber, an engineer or maybe even a physicist. Consider next the link from actual water temperature to the temperature gap. Algebraically the gap is defined as the difference between the desired and actual water temperature (temperature gap = desired water temperature – actual water temperature). But a meaningful temperature gap in a shower also requires a process for sensing the gap. The existence of a temperature
gap alone does not guarantee goal-seeking behaviour. For example, if someone entered a shower in a winter wetsuit, complete with rubber hood and boots, they would not notice a temperature gap, and the entire feedback loop would be rendered inactive. Although this case is extreme and fanciful, it illustrates the importance of credibly grounding causal links.

The final link in the balancing loop is from temperature gap to the flow of hot water. Arguably this is the single most important link in the loop because it embodies the decisionmaking process for adjusting the flow of hot water. There is a huge leap of causality in this part of the diagram. The commonsense interpretation of the link is that when any normal person feels too hot or too cold in a shower, he or she will take corrective action by adjusting the flow of hot water. But how do they judge the right amount of corrective action? How quickly do they react to a temperature gap and how fast do they turn the tap? All these factors require consideration. Moreover, the key to over-reaction in showers arguably lies in this single step of causality. Why do people get trapped into a repetitive hot-cold cycle when all they normally want to achieve is a steady comfortable temperature? The answer must lie in how they choose to adjust the tap setting, in other words in their own decision-making process.

2.2.1.3 Simulation of a Shower and the Dynamics of Balancing Loops

Figure 2.8 shows the simulated dynamics of a slow-to-respond shower over a period of 120s generated by a simulation model containing all the processes mentioned above. As before the desired water temperature is a cool 25°C. However in this scenario the water temperature starts too high at 40°. Corrective action lowers the temperature at the shower-head to the desired 25° in about 10 s, but the temperature continues to fall, reaching a minimum just below 24° after 12s. Further corrective action then increases the temperature, leading to an overshoot

![Fig. 2.8 Simulated dynamics of a slow-to-respond shower](image-url)
that peaks at 27° after 21s. The cycle repeats itself twice in the interval up to 60 s, but each time the size of the temperature overshoot and undershoot is reduced as the showertaker gradually finds exactly the right tap setting for comfort. In the remainder of the simulation, from 60 to 120s, the temperature at the shower-head remains steady at 25°. The overall trajectory is a typical example of goal-seeking dynamics arising from a balancing loop with delay.

It is worthwhile to remember this particular combination of feedback structure and dynamic behaviour because balancing loops crop up all over the place in business, social, environmental and biological systems. Wherever people, organisations or even organisms direct their efforts and energy to achieving and maintaining specific goals in the face of an uncertain and changing environment there are balancing loops at work. Companies set themselves sales objectives, quality standards, financial targets and goals for on-time delivery. Governments set targets for economic growth, inflation, hospital waiting times, literacy, exam pass rates, road congestion, and public transport usage. The human body maintains weight, balance, temperature, and blood sugar. The ecosystem sustains an atmosphere suitable for the animals and plants within it. The vast global oil industry maintains a supply of oil sufficient to reliably fill our petrol tanks. The electricity industry supplies just enough electricity to keep the lights on. Economies generate enough jobs to keep most people employed. The list goes on and on. In some cases, like people’s body temperature or domestic electricity supply, the balancing process works so well that it is rare to find deviations from the ‘goal’ – a degree or two from normal body temperature is a sign of illness and, in the electricity industry, it is unusual (at least in the developed world) for the lights to dim. In many cases, like sales objectives or hospital waiting times, the goals are known, but performance falls chronically short or else gently overshoots and undershoots. But in other cases, like employment in the economy or inventory levels in supply chains, the balancing process is far from perfect. Performance deviates a long way from the goal, too much or too little. Corrective action leads to over and under compensation and the goal is never really achieved, at least not for long.

2.2.2 From Events to Dynamics and Feedback Structure: Drug Related Crime

A shift of mind (from event oriented thinking to feedback systems thinking) is not easy to achieve. The best way to make progress is through examples of feedback systems thinking applied to real-world situations. So, instead of hot water showers we now consider something entirely different – drug related crime. A typical description of the problem, by the victims of crime, might be as follows.

Drugs are a big worry for me, not least because of the crimes that addicts commit to fund their dependency. We want the police to bust these rings and destroy the drugs. They say they’re doing it and they keep showing us sacks of cocaine that they’ve seized, but the crime problem seems to be getting worse.
Expressed this way drug related crime appears as a series of disturbing events. There is a concern about crime among the members of the community affected by it. They want action backed-up with evidence of police attempts to fix the problem by busting rings and seizing drugs. But, despite these efforts, more crimes are happening. The feedback systems thinker re-interprets the description and draws out those aspects concerned with performance through time (dynamics) that suggest an underlying feedback structure, one or more interacting feedback loops, capable of generating the dynamics of interest. Of particular significance are *puzzling* dynamics, performance through time that people experience but do not want or intend. Some of the most interesting and intractable problems in society and business appear this way.

Figure 2.9 shows the unintended dynamics of drug related crime that might be inferred from the brief verbal description above. This is just a rough sketch to provide a focus for structuring the problem. On the horizontal axis is time in years. On the vertical axis is drug related crime defined in terms of ‘incidents per month’. There are two trajectories. The upper line is a sketch of crime reported by the community. We assume a growth trajectory because ‘the crime problem seems to be getting worse’. The lower line is a sketch of tolerable crime, a kind of benchmark against which to compare the actual level of crime. We assume a downward sloping trajectory because the community wants less crime and fewer drugs, and the police are taking action to achieve this end by seizing drugs and arresting dealers.\(^4\)

The divergence between reported and tolerable crime is of particular interest to the feedback systems thinker. What feedback structure could explain this phenomenon? Reported crime is growing and we know that growth arises from reinforcing feedback.

![Fig. 2.9 Unintended dynamics of drug related crime – a rough sketch](image)

\(^4\)You may be thinking this method of creating time charts is rather loose and in a sense you are right because we have very little data about the problem. But even in practice, with real clients, the information sources for modelling are always a pragmatic blend of informed opinion, anecdote, objective facts and clear reasoning. For a good example of this balanced approach in the area of drug policy, see Homer (1993) and Levin et al. (1975).
So where could such a malignant feedback process come from and why would it exist at all if those involved want less crime, not more? The persistence of unwanted growth in crime suggests a feedback loop that weaves its way around society (crossing the boundaries between police, the community and drug users) and by doing so it goes unnoticed.

2.2.2.1 Feedback Loops in Drug Related Crime

Figure 2.10 is a causal loop diagram for drug related crime. First consider the words and phrases alone. They provide the basic vocabulary of the causal model, the factors that drive-up crime, or at least are hypothesised to do so. They also give clues to the boundary of the model, which parts of society are included. Of course there is drug related crime itself, the variable of central interest and concern to the community. There is a ‘call for police action’ and drug seizures that take us inside the police department. Then there is supply, demand and price that belong in the world of drug users who commit crime.5

These factors join-up to make a closed loop of cause and effect. The loop brings together disparate parts of society to reveal a surprise. Hidden in the connections is a reinforcing feedback process responsible for (or at least contributing to) escalating crime. To confirm let’s trace the effect around the loop of an imagined increase in drug related crime. In this kind of analysis the reason for the initial increase does not matter, it is the feedback effect that is of central interest. The story begins at the

---

5Notice that all the terms in the diagram are nouns or so-called ‘noun-phrases’. This is an important diagramming convention because you want concepts to denote things, attributes or qualities that can, in imagination, be unambiguously increased or decreased. Then, and only then, is it possible to cleanly assign polarity to causal links and thereby deduce the loop types – balancing or reinforcing. Take for example price and drug related crime. It is easy to imagine the price of drugs going up or down and separately to imagine drug related crime increasing or decreasing. Therefore, when a causal link is drawn between these two concepts, it is meaningful to ask whether an increase in one leads to an increase or decrease in the other. This thought experiment would make no sense if one or other concept were labelled as an activity, say pricing instead of price.
top of the diagram. An increase of drug related crime leads to a call for more police action. More police action (raids and arrests) leads to more drug seizures. So far so good. But the paradox lies in what happens next as available drugs are traded on the streets. An increase in drug seizures causes the supply of drugs to decrease. This supply cut then causes the price of drugs to increase, just like any traded goods subject to market forces, assuming of course that higher price does not depress demand. And crucially for illegal drugs, price has little effect on demand because most users are addicts, dependent on their daily fix. So an increase in price merely boosts crime as desperate drug users steal even more to fund their addiction. The reinforcing loop is plain to see. There is a ‘crime spiral’ in which any increase of drug related crime tends to amplify itself through the inadvertent actions of police, drug dealers and addicts.

2.2.2.2 Scope and Boundary of Factors in Drug Related Crime

There could be more, much more, to the problem situation than the six concepts shown. So I am not saying these six factors and this single reinforcing loop is a perfect representation of escalating crime in a community plagued with drug addicts. Rather it is a useful way of thinking about the problem that raises the perspective above the narrow confines of a single stakeholder. In fact three stakeholders are united in this particular view. And, just as we noted in the shower case, there is a lot going on behind the scenes of the stark causal links; detail that would need to be fleshed out in thinking more carefully about the problem and in building a simulation model to test alternative intervention policies. There is the community suffering from crime and calling for police action. There is the police department, concerned with all sorts of law enforcement, allocating police officers to priority tasks, among which is drug busting. And then there is the shady world of drug dealers sourcing drugs and covertly selling them to addicts who must consume, no matter what the cost. Later in the chapter we see how this qualitative feedback diagram is transformed into a full-blown simulator. But for now I want to end the discussion of drug related crime by inviting you to think about what else might be included in a conceptual model of the problem.

One place to expand the diagram is with demand and supply. (Another good idea in practice is to gather more time series data to help refine the dynamic hypothesis, but we will by-pass that step in this small illustrative example.) What if there is growth in demand because addicts and dealers themselves recruit new users? This possibility adds a whole new dimension to escalating crime not dealt with in our current picture, a new theory if you like. What if, as is surely the case, the available supply of drugs increases as the price rises? Does that mean drug seizures perversely expand the whole illegal drug industry (in the long run) by artificially boosting prices? Such industry growth could exacerbate the crime problem, particularly if the relevant time frame is a decade or more rather than just a few years. These questions, and others like them, are worth probing and may usefully expand the scope and boundary of our thinking. The point however, in any such conceptualisation task, is to avoid unnecessary complexity and focus on
finding plausible loops, often unnoticed in the pressure of day-to-day operations, that not only challenge conventional event-oriented thinking but also produce dynamics consistent with the observed problem.

2.2.2.3 An Aside: More Practice with Link Polarity and Loop Types

I have explained the origin of the reinforcing loop in Fig. 2.10 by tracing an imagined change in crime all the way around the loop and showing it leads to even more crime. As mentioned earlier, another way to find the loop type is to use the counting rule. Count the negative links around the loop. If the number of links is odd then the loop is balancing and if the number is even the loop is reinforcing. Let’s do this exercise now. First we need to assign link polarities using the standard test. Any individual link connects two concepts A and B where A is the cause and B is the effect. For each link imagine an increase in the cause A and then work out the effect on B. In this thought experiment all other influences on B are assumed to remain unchanged, the ceteris paribus assumption. The link is positive if, when A increases, B increases above what it would otherwise have been. The link is negative if, when A increases, B decreases below what it would have been. Note that the mirror image test works too. So when A decreases and B also decreases the link is positive, but when A decreases and B increases the link is negative. What matters for polarity is whether or not there is a reversal.

We start at the top. All else equal, if drug related crime increases then the call for police action (complaints from the community) increases above what it would otherwise have been, a positive link. When the call for police action increases then drug seizures increase, another positive link. Note there is a large leap of causality here that relies on all else remaining equal, ceteris paribus. We implicitly assume that a call for action really leads to action (in this case more police allocated to drug busting), rather than being ignored. Moreover, we assume that more police leads to more seizures. In the next link an increase in seizures leads to a decrease in supply, below what it would otherwise have been, a negative link. Then a decrease in supply leads to an increase in price, another negative link coming this time from a mirror image test. Here there is a particularly clear instance of ceteris paribus reasoning because price depends both on supply and demand. The assumption behind the polarity test is that demand remains constant. An equivalent test on the demand-to-price link shows it is positive: an increase in demand leads to an increase in price, assuming supply is held constant. Finally an increase in price leads to an increase in drug related crime, a positive link that completes the loop. Counting-up there are two negative links around the loop, an even number, so the loop type is reinforcing.

2.2.2.4 Purpose and Use of Causal Loop Diagrams: A Summary

As we have seen, causal loop diagrams offer a special overview of business and society, showing what is connected to what and how changes in one part of the system might propagate to others and return. People often say we live in an
interconnected world. But we have no way, other than words, to express this complexity. Causal loop diagrams, concise and visual, reveal the interconnections, both obvious and hidden. Moreover, they can be used to elicit and capture the mental models of individuals or teams and to expand the boundary of people’s thinking beyond the parochial.

Causal loop diagrams also capture hypotheses about dynamic behaviour. Here is the beginning of the shift of mind so vital to feedback systems thinking. The future time path of any organisation is partly and significantly pre-determined by its structure, the network of balancing and reinforcing feedback loops that drive performance through time. Causal loop diagrams embody this important philosophical view by making plain the important feedback loops believed to be responsible for observed performance.

2.2.2.5 Basic Tips: Picking and Naming Variables

The choice of words is vital. Each variable must be a noun. Avoid the use of verbs or directional adjectives. For example a causal diagram can use the word ‘sales’ but not ‘sales planning’ or ‘increased sales’. Simple nouns like ‘accounts’ or ‘staff’ can be augmented with adjectives to give phrases like ‘large accounts’ or ‘experienced staff’. Sticking to these simple naming rules helps when assigning polarity to causal links and explaining how changes propagate around loops.

Words are versatile, but they should also be grounded in facts. The range of concepts that can be included in causal loop diagrams extends from the hard and easily measureable, such as ‘new products’ and ‘recruits’, to the soft and intangible such as ‘morale’ or ‘customer perceived quality’. A powerful feature of feedback systems thinking and system dynamics is its ability to incorporate both tangible and intangible factors. However, for any variable no matter how soft, you should always have in mind a specific unit of measure, a way in which the variable might be quantified, even if formal recorded data do not exist. So you might imagine morale on a scale from 0 (low) to 1 (high) or product quality on a scale from 1 (low) to 5 (high). And be sure to pick words that imply measureability, such as ‘delivery lead time’ thought of in weeks or months, rather than a vague concept like ‘delivery performance’.

2.2.2.6 Basic Tips: Meaning of Arrows and Link Polarity

Arrows show the influence of one variable on another – a change in the cause leads to a change in the effect. The assignment of link polarity (+) or (−) makes the direction of change clear. So in Fig. 2.11 an increase in marketing budget leads to an increase in sales, which is a positive link.

Polarity assignment works equally well for intangible variables such as industry reputation in the lower half of Fig. 2.11. Industry reputation here is an intangible concept measured on a scale from 0 to 1. An increase in industry reputation leads to an increase in customers interested.
A useful refinement in polarity assignment is to note whether the effect of a given change is an increase (or decrease) more than it would otherwise have been. The use of this extra phrase avoids ambiguity in situations where the effect is cumulative. For example customers are likely to be accumulating over time and therefore the effect of rising industry reputation is to attract more customers than there would otherwise have been.

### 2.2.2.7 Basic Tips: Drawing, Identifying and Naming Feedback Loops

For the systems thinker, feedback loops are the equivalent of the sketches created by political cartoonists. They capture something important about the situation or object of interest. Just as a few bold pen lines on a canvas can characterise George Bush, Osama bin Laden, or Margaret Thatcher, so a few feedback loops on a whiteboard can characterise an organisation. Like celebrity sketches feedback loops should be drawn clearly to identify the dominant features, in this case important loops. Sterman (2000) identifies five tips for visual layout:

1. Use curved lines to help the reader visualise the feedback loops
2. Make important loops follow circular or oval paths
3. Organise diagrams to minimise crossed lines
4. Don’t put circles, hexagons, or other symbols around the variables in causal diagrams. Symbols without meaning are ‘chart junk’ and serve only to clutter and distract.
5. Iterate. Since you often won’t know what all the variables and loops will be when you start, you will have to redraw your diagrams, often many times, to find the best layout.

As we have already seen, for the hot water shower and drug related crime, there are two main loop types, balancing and reinforcing. A loop type is identified by imagining the effect of a change as it propagates link-by-link around the loop.
A reinforcing loop is one where an increase in a variable, when traced around the loop, leads to a further increase in itself. Such an outcome requires an even number (or zero) of negative links. A balancing loop is one where an increase in a variable, when traced around the loop, leads to a counterbalancing decrease in itself. Such an outcome requires an odd number of negative links. Once you have identified loop types it is good practice to label them $R$ for reinforcing and $B$ for balancing, the letter encircled by a small curved arrow drawn clockwise for clockwise loops (and vice versa).

By following these tips and by studying the examples in the chapter you should be able to create, label and interpret your own causal loop diagrams. Often you will end up with multiple interlocking loops that reach across conventional organisational boundaries. Then it is particularly important to follow the five tips for visual layout mentioned above. A good example can be found later, in Part 2.2.6, based on a study of orchestra management. Selected feedback loops show that factors affecting the success of an orchestra reach well beyond the boundaries of the concert hall.

### 2.2.3 Modelling to Simulate Dynamic Systems

Causal loops diagrams are very effective for expanding the boundary of your thinking and for communicating important interdependencies. But they are not especially good as the basis for a full-blown model and simulator that computes dynamics and performance through time. For a working model we need better resolution of the causal network. It turns out there is more to causality and dynamics than words and arrows alone. The main new concepts required to make simulators are introduced in this section. They transform a simple sketch of causality into a portrait (or better still an animation) that brings feedback loops to life, by specifying the realistic processes that lie behind causal links as the basis for an algebraic model and simulator.

#### 2.2.3.1 Asset Stock Accumulation

Asset stock accumulation is a very important idea in system dynamics, every bit as fundamental as feedback and in fact complementary to it. You can’t have one without the other. Asset stocks accumulate change. They are a kind of memory, storing the results of past actions. When, in a feedback process, past decisions and actions come back to influence present decisions and actions they do so through asset stocks. Past investment accumulates in capital stock – the number of planes owned by an airline, the number of stores in a supermarket chain, the number of ships in a fishing fleet. Past hiring accumulates as employees – nurses in a hospital, operators in a call centre, players in a football squad, faculty in a university. Past production accumulates in inventory and past sales accumulate in an installed base. All business and social systems contain a host of different asset stocks or resources.
that, when harnessed in an organisation, deliver its products and services. And, crucially, the performance over time of an enterprise depends on the balance of these assets and resources (Warren 2008). An airline with lots of planes and few passengers is out of balance and unprofitable. Empty seats bring no revenue. A factory bulging with inventory while machines lie idle is out of balance and underperforming. Inventory is expensive.

To appreciate how such imbalances occur we first need to understand the nature of asset stock accumulation – how assets build and decay through time. A process of accumulation is not the same as a causal link. Accumulations change according to their inflows and outflows in just the same way that water accumulates in a bathtub. If the inflow is greater than the outflow then the level gradually rises. If the outflow is greater than the inflow then the level gradually falls. If the inflow and outflow are identical then the level remains constant. This bathtub feature of assets in organisations is depicted using the symbols in Fig. 2.12. Here an asset stock or resource is shown as a rectangle, partially filled. On the left is an inflow comprising a valve or tap superimposed on an arrow. The arrow enters the stock and originates from a source, shown as a cloud or pool. A similar combination of symbols on the right represents an outflow. In this case the flow originates in the stock and ends up in a sink (another cloud or pool). The complete picture is called a stock and flow network.

Consider for example a simple network for university faculty as shown in Fig. 2.13. Let’s forget about the distinction between professors, senior lecturers and junior lecturers and call them all instructors. Instructors teach, write and do research. The stock in this case is the total number of instructors. The inflow is the rate of recruitment of new faculty – measured say in instructors per month, and the outflow is turnover – also measured in instructors per month. The source and sink represent the university labour market, the national or international pool of academics from which faculty are hired and to which they return when they leave. The total number of instructors in a university ultimately depends on all
sorts of factors such as location, reputation, funding, demand for higher education and so on. But the way these factors exert their influence is through flow rates. Asset stocks cannot be adjusted *instantaneously* no matter how great the organisational pressures. Change takes place only gradually through flow rates. This vital *inertial* characteristic of stock and flow networks distinguishes them from simple causal links.

### 2.2.3.2 Accumulating a ‘Stock’ of Faculty at Greenfield University

The best way to appreciate the functioning of stocks and flows is through simulation. Luckily it is only a small step from a diagram like Fig. 2.13 to a simulator. Use this URL [http://www.iseesystems.com/community/downloads/OpenUniversity.aspx](http://www.iseesystems.com/community/downloads/OpenUniversity.aspx) to find the model called Stock Accumulation – Faculty and open it. A stock and flow network just like Fig. 2.13 will appear on the screen. To make this little network run each variable must be plausibly quantified. Imagine a new university called Greenfield. There is a small campus with some pleasant buildings and grounds, but as yet no faculty. The model is parameterised to fit this situation. Move the cursor over the stock of instructors. The number zero appears meaning there are no instructors at the start of the simulation. They will come from the academic labour market. Next move the cursor over the valve symbol for recruitment. The number five appears. This is the number of new instructors the Vice Chancellor and Governors plan to hire each month. Finally move the cursor over the symbol for turnover. The number is zero. Faculty are expected to like the university and to stay once they join. So now there is all the numerical data to make a simulation: the starting size of the faculty (zero), intended recruitment (five per month) and expected turnover (zero per month).

Press the Run button. What you see is stock accumulation as the ‘bathtub’ of instructors gradually fills-up. This steady increase is exactly what you expect if, each month, new instructors are hired and nobody leaves. Now double click on the graph icon. A chart appears, just like Fig. 2.14, that plots the numerical values through time of instructors (line 1), recruitment (line 2) and turnover (line 3). The horizontal time axis spans 12 months. The number of instructors begins at 0 and builds steadily to 60 after 12 simulated months. Meanwhile recruitment remains steady at five instructors per month and turnover is zero throughout. Numerically the simulation is correct and internally consistent. Recruitment at a rate of 5 instructors per month for 12 months will, if no-one leaves, result in a faculty of 60 people.

That’s all very obvious, and in a sense, stock accumulation is no mystery. It is simply the result of taking the numerical difference, period by period, between the inflow and the outflow and adding it to the stock size. An equation shows the simple arithmetic involved:

\[
\text{instructors} (t) = \text{instructors} (t – dt) + (\text{recruitment} – \text{turnover}) \times dt
\]

\[
\text{INIT instructors} = 0
\]
Here the number of instructors at time t (this month) is equal to the number of instructors at time t-dt (last month) plus the difference between recruitment and turnover over an interval of time dt. The interval is a slice of time convenient for the calculation, the so-called delta-time dt. So if dt is equal to 1 month then the calculation is a monthly tally of faculty. The initial value of instructors is set at zero.

All stock accumulations have the same mathematical form, no matter whether they represent tangible assets (machines, people, planes) or intangible assets (reputation, morale, perceived quality). The relationship between a stock and its flows is cumulative and naturally involves time. It is not the same as a causal link. The stock of instructors accumulates the net amount of recruitment and turnover through time. Mathematically speaking the stock integrates its inflow and outflow. The process is simple to express, but the consequences are often surprising.

To illustrate let’s investigate a 36 month scenario for Greenfield University. Recruitment holds steady at five instructors per month throughout, but after 12 months some faculty are disillusioned and begin to leave. To see just how many leave double click on the turnover icon. A chart and table appear. The chart on the left shows the pattern of turnover across 36 months and the table on the right shows the corresponding numerical values at intervals of 3 months. For a period of 12 months turnover is zero and faculty are content. Then people start to leave, at an increasing rate. By month 15 turnover is two instructors per month, by month 18 it is four instructors per month, and by month 21 it is six instructors per month. The upward trend continues to month 27 by which time faculty are leaving at a rate of ten per month. Thereafter turnover settles and remains steady at ten instructors per month until month 36. (As an aside it is worth noting this chart is a just an assumption about future turnover regardless of the underlying cause. In reality instructors may leave Greenfield University due to low pay,
excess workload, lazy students, etc. Such endogenous factors would be included in a complete feedback model.)

To investigate this new situation it is first necessary to extend the simulation to 36 months. Close the turnover chart by clicking the OK button. Then find Run Specs in the pull-down menu called Run at the top of the screen. A window appears containing all kinds of technical information about the simulation. In the top left there are two boxes to specify the length of simulation. Currently the simulator is set to run from 0 to 12 months. Change the final month from 12 to 36 and click OK. You are ready to simulate. However, before proceeding, first sketch on a blank sheet of paper the faculty trajectory you expect to see. A rough sketch is fine – it is simply a benchmark against which to compare model simulations. Now click the run button. You will see the ‘bathtub’ of faculty fill right to the top and then begin to empty, ending about one quarter full. If you watch the animation very carefully you will also see movement in the dial for turnover. The dial is like a speedometer, it signifies the speed or rate of outflow. Now move the cursor over the turnover icon. A miniature time chart appears showing the assumed pattern of turnover. Move the cursor over recruitment and another miniature time chart appears showing the assumed steady inflow of new faculty from hiring. Finally move the cursor over the stock of instructors. The time chart shows the calculated trajectory of faculty resulting from the accumulation of recruitment (the inflow) net of turnover (the outflow).

All three trajectories can be seen in more detail by clicking the graph icon. The chart in Fig. 2.15 appears. Study the time path of instructors (line 1). How does the shape compare with your sketch? For 12 months the number of instructors grows in a straight line, a simple summation of steady recruitment (line 2). Then turnover begins to rise (line 3). The faculty therefore grows less quickly.

![Fig. 2.15 Faculty size at Greenfield University – a 36 month simulation](image-url)

Fig. 2.15 Faculty size at Greenfield University – a 36 month simulation
By month 20 turnover reaches five instructors per month, exactly equal to recruitment, and line 3 crosses line 2. The process of accumulation is perfectly balanced. New faculty are arriving at the same rate existing faculty are leaving. The number of instructors therefore reaches a peak. Beyond month 20 turnover exceeds recruitment and continues to rise until month 27 when it reaches a rate of ten instructors per month, twice the recruitment rate. The faculty shrinks even though turnover itself stabilises.

Notice that although the number of instructors gently rises and falls, neither the inflow nor the outflow follow a similar pattern. The lack of obvious visual correlation between a stock and its flows is characteristic of stock accumulation and a clear sign that the process is conceptually different from a causal link. You can experience more such mysteries of accumulation by redrawing the turnover graph and re-simulating. Double click on the turnover icon and then hold down the mouse button as you drag the pointer across the surface of the graph. A new line appears and accordingly the numbers change in the table on the right. With some fine-tuning you can create a whole array of smooth and plausible turnover trajectories to help develop your understanding of the dynamics of accumulation. One interesting example is a pattern identical to the original but scaled down, so the maximum turnover is no more than five instructors per month.

2.2.3.3 The Coordinating Network

Feedback loops are formed when stock and flow networks interact through causal links, in other words when the inflows and outflows of one asset stock depend, directly or indirectly, on the state or size of other asset stocks. In principle all the stocks and flows in an organisation are mutually dependent because conditions in one area or function may cause or require changes elsewhere. Coordination is achieved through a network that relays the effect, direct or indirect, of particular stocks on a given flow. The symbols used for the coordinating network are shown on the left of Fig. 2.16. A causal link is drawn as an arrow with a solid line, exactly
the same as in a causal loop diagram. An information flow is drawn as an arrow with a dotted line. It too depicts an influence of one variable on another though in a subtly different way.

A converter represents a process that converts inputs into an output and is depicted as a circle. Converters receive causal links or information flows and transform them according to whatever rules, physical laws or operating policies apply. In a simulator there is an equation behind each converter that specifies the rules, as we will see shortly.

The diagram on the right of Fig. 2.16 shows how all the symbols fit together. There are two stock and flow networks joined by a coordinating network containing three converters. This is a feedback representation because the two flow rates not only accumulate into the two stocks but are themselves regulated by the magnitude of the stocks. The picture can readily be extended from 2 to 20 stocks or more depending on the complexity of the situation at hand. No matter how large the picture, it captures an elaborate process of bootstrapping that arises from nothing more than cause, effect, influence and accumulation found in all organisations.

### 2.2.3.4 Modelling Symbols in Use: A Closer Look at Drug Related Crime

To see all the modelling symbols in use we revisit the problem of drug related crime. Recall the original intention was to identify systemic factors that explain growth in drug related crime despite the drug busting efforts of police. Figure 2.17 shows the sectors of society involved and one important feedback loop, a reinforcing crime spiral. There are four sectors: the community itself (suffering from crime), the police department (trying to control crime), the street market for drugs, and the world of the drug users.
market for drugs and the world of the drug user. A simulatable model of this situation represents the stock accumulations, causal links, information flows and operating policies that lie behind the reinforcing crime spiral. The model is presented sector by sector.

Figure 2.18 shows the causal links in the community. The community is concerned about drug related crime and raises its collective concern through a call for police action. Notice that each concept is accompanied by units of measure that help ground the model and subsequently aid quantification. The search for practical and consistent units of measure is an important modelling and thinking discipline. Drug related crime is expressed as incidents per month. A practical measure of the community’s ‘call for police action’ is complaints per month. The link here is the same as in the causal loop diagram. But the difference in units between the cause and effect shows the need for another concept, community sensitivity to crime, to operationalise the original link. Community sensitivity can be thought of in terms of complaints per incident. A community that is very sensitive to crime will generate more complaints per incident than a community resigned or indifferent to crime, thereby bringing to bear more pressure for police action.

Figure 2.19 takes us inside the police department. Notice that the police department converts the call for police action (in complaints per month) into drug seizures (in kg per month). In the causal loop diagram this conversion of complaints into seizures is achieved in a single causal link. The stock-and-flow diagram reveals the operating detail behind the link. In the middle of the diagram there is a stock accumulation representing the number of police allocated to drug busting.

The policy controlling the allocation of police is in the top half of the diagram and is a typical goal-seeking adjustment process. Call for police action leads to an indicated allocation of police – the number of police officers deemed necessary to deal with the drug problem. This goal is implemented by reallocating police between duties. The change in allocation of police (measured in police officers per month) depends on the difference between the indicated allocation, the current number of police allocated to drug busting and the time it takes to move staff.
In reality the process of reallocating police takes time and organisational effort, all of which is captured by the stock and flow network for number of police.

Incidentally, the ‘cloud’ on the left of this network represents the total pool of police in the department, currently working on other duties, who might be called into drug busting.\(^6\) The amount of drug seizures is proportional to the number of police allocated. To operationalise this link it is necessary to introduce a new concept ‘police effectiveness in drug busting’ measured in kilograms per officer per month – a kind of drug busting productivity.

The street market for drugs adjusts the street price of drugs according to the supply and demand of drugs, as shown in Fig. 2.20. The supply of drugs on the street is equal to the total supply of drugs less drug seizures. The drug supply gap is the difference between demand for drugs and supply on the street (all measured in kilograms per month). The existence of a supply gap generates pressure for price change which in turn drives the change in street price that accumulates in the street price (measured in £ per kilogram). The pricing ‘policy’ here is informal – an invisible hand. Note there is no target price. The price level continues to change as long as there is a difference between supply and demand.

In Fig. 2.21 we enter the world of drug-dependent users with an addiction and craving that must be satisfied at all costs, even if it involves crime. Addicts need funds (in £ per month) to satisfy their addiction. In a given geographical region the

\(^6\)By using a cloud symbol we assume that the pool of police officers assigned to duties other than drug busting is outside the boundary of the model. If for some reason we wanted to track the number of officers in this pool, then the cloud symbol would be replaced by a stock accumulation with its own initial number of officers. It would then be apparent from the diagram that shifting more officers to drug-busting reduces the number available to work on other duties.
funds required by addicts are proportional to their collective demand for drugs (in kilograms per month) and the prevailing street price (in £ per kilogram). Drug related crime is the amount of crime (in incidents per month) necessary to raise the funds required. This conversion of funds into crime depends also on the average yield per crime incident (measured in £ per incident), which is a measure of criminal productivity and reflects the wealth of the burgled community.
2.2.4 Equation Formulations

The final step in developing a simulator is to write algebraic equations. Full structural diagrams are a good starting point because they show all the variables that must appear in the equations. Nevertheless there is skill in writing good algebra in a way that properly captures the meaning of the relationships depicted.

2.2.4.1 Drug Related Crime

Consider the formulation of drug related crime. We know from the diagram that drug related crime depends on the funds required (by addicts) to satisfy their addiction and on the average yield per crime incident. These two influences are reproduced in Fig. 2.22. But how are they combined in an equation? Should they be added, subtracted, multiplied or divided? The top half of Fig. 2.22 is a plausible formulation where drug related crime is equal to funds required divided by average yield. This ratio makes sense. We would expect that if addicts require more funds they will either commit more crimes or else operate in a neighbourhood where the yield from each crime is greater. So funds required appears in the numerator and average yield in the denominator. The ratio expresses precisely and mathematically what we have in mind.

An alternative formulation, such as the product of ‘funds required’ and ‘average yield’, contradicts commonsense and logic. A simple numerical example shows just how ludicrous such a multiplicative formulation would be. Let’s suppose there are ten addicts in a neighbourhood and collectively they require £1,000 per month to satisfy their addiction. If each member contributes £100 per month then the total funds required is £1,000 per month. But if the average yield per crime incident is £50 then the total drug related crime is £20,000 per month. This is ridiculous! The correct formulation is to divide the funds required by the average yield per crime incident.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Drug Related Crime [incidents/month] =</td>
<td>Funds Required to Satisfy Addiction [£/month] / Average Yield Per Crime Incident [£/incident]</td>
<td></td>
</tr>
</tbody>
</table>

Dimensional Analysis

Left hand side: [incidents/month]
Right hand side: [£/month]/[£/incident] = [£/month]*[incident/£] = [incidents/month]

Fig. 2.22 Equation formulation for drug related crime

Drug Related Crime

Funds Required to Satisfy Addiction

Average Yield Per Crime Incident

£ per incident

incidents per month

£ per month
satisfy their addiction. On average each crime incident yields £100. A multiplicative formulation would imply that drug related crime in the neighbourhood takes a value of \((1,000 \times 100)\), in other words one hundred thousand – which is numerically implausible and wrong. The correct formulation results in \((1,000/100)\), or ten incidents per month.

There are numerous guidelines for equation formulation to help modellers write good algebra that means what they intend. One of the most useful is to ensure dimensional consistency among the units of measure in an equation. This guideline is always useful in situations like the one above where the main formulation challenge is to pick the right arithmetical operation. Dimensional consistency requires that the units of measure on left and right of an equation match. In this case ‘drug related crime’ on the left is measured in incidents per month. So the operation on the right must combine ‘funds required’ and ‘average yield’ in such a way as to create incidents per month. Taking the ratio of funds required [£/month] to average yield [£/incident] achieves this outcome, as shown in the dimensional analysis box of Fig. 2.22. No other simple arithmetic operator such as +, −, or * leads to this result. For example the units of measure for a multiplicative formulation would be £² per month per incident, a bizarre and meaningless metric that reveals a fatal formulation error.

2.2.4.2 Funds Required to Satisfy Addiction

The formulation for funds required is shown in Fig. 2.23. We know from the diagram that ‘funds required’ depends on demand for drugs and the street price. The greater is demand, or the higher the street price, the more funds required. Moreover, a combination of greater demand and higher street price calls for even more funds and therefore suggests a multiplicative formulation. So the equation for ‘funds required’ is expressed as the product of demand for drugs and the street price. A dimensional analysis shows the units of measure are consistent in this formulation.

2.2.4.3 Street Price and Price Change

The street price of drugs is a stock that accumulates price changes. The change in street price is a function of street price itself and ‘pressure for price change’. This pressure depends on the drug supply gap, in other words whether there is an adequate supply of drugs on the street. The diagram and corresponding equations are shown in Fig. 2.24. The first equation is a standard formulation for a stock accumulation. The street price this month is equal to the price last month plus the change in price during the month.

The change in street price arises from informal, covert trading of illegal drugs on street corners. It is an important formulation that depends both on street price itself and the pressure for price change. This pressure is itself a function of the drug supply
gap, a graphical function whose shape is sketched in the lower left of Fig. 2.24. To understand the price change formulation, first imagine the drug supply gap is zero – so there is just enough volume of drugs being supplied by dealers to satisfy demand. Under this special condition the pressure for price change is logically zero and so too is the change in street price itself. The multiplicative formulation ensures
no price change when the pressure for price change is zero. Now suppose there is a shortage of drugs on the street. The drug supply gap is positive and, through the graph, the pressure for price change is also positive. Moreover, as the gap grows the pressure rises more quickly than a simple linear proportion. The relationship is non-linear, with increasing gradient. A mirror image applies when there is a surplus of drugs and the drug supply gap is negative. Pressure for price change is expressed as a fraction per month, so the resulting change in price is the street price itself multiplied by this fraction per month. The units of price change are £/kilogram/month and an inspection of the price change equation shows the required dimensional balance.

Notice that the price itself feeds back to influence price change. This is quite a subtle dynamic formulation and has the curious, though realistic, implication that there is no pre-determined market price or cost-plus anchor toward which price adjusts. The only meaningful anchor is the current price. So if there is a chronic undersupply the price will relentlessly escalate, and conversely if there is a chronic oversupply the price will steadily fall. Price settles at whatever level it attains when supply and demand are balanced, no matter how high or low.

### 2.2.4.4 Allocation of Police

The formulation for the allocation of police is shown in Fig. 2.25. It is a classic example of an asset stock adjustment process. At the heart of the formulation is a stock accumulation of police officers guided by an operating policy for adjusting the allocation of police. The first equation is a standard stock accumulation in
which the number of police allocated to drug busting this month is equal to the number allocated last month plus the change in allocation during the month. The second equation represents the policy for redeploying police to drug busting. The change in allocation of police depends on the gap between the indicated allocation of police and the current number of police allocated to drug busting. If there is pressure from the community on the police department to deal with crime then this gap will be positive and measures how many more police officers are really needed. However officers are redeployed gradually with a sense of urgency captured in the concept ‘time to move staff’. The greater this time constant, the slower the rate of redeployment for any given shortfall of police officers. The formulation divides the shortfall by the time to move staff, resulting in a dimensionally balanced equation with appropriate units of police officers per month.

2.2.5  System Dynamics in Action: The Rise of Low-Cost Air Travel in Europe

A modelling project is iterative. It begins with a concern about dynamics (performance over time) in the real world and preliminary ideas about feedback structure. Then gradually, in stages, a model takes shape that clarifies the concern and sharpens ideas about structure. But the purpose is not to create a perfect model that replicates the real world situation in every detail. Rather it is to use modelling as a learning process to investigate, discover and clarify feedback structure and the dynamic behaviour it implies.

To illustrate we examine a model about the early growth strategy of easyJet, one of the UK’s most successful no-frills airlines, at the dawn of low-cost flights in Europe. To appreciate the model’s boundary and scope it is important to imagine the European airline industry not as it is today but as it was back in the mid-1990s when full-service air travel was the norm and low cost flights were a new and unproven business concept.

2.2.5.1  easyJet: A Bright Idea, But Will It Work?

The historical situation is described in an article called ‘easyJet’s $500 Million Gamble’ (Sull 1999). The opening paragraph sets the scene.

This case study details the rapid growth of easyJet which started operations in November 1995 from London’s Luton airport. In 2 years, it was widely regarded as the model low-cost European airline and a strong competitor to flag carriers. The company has clearly identifiable operational and marketing characteristics, e.g. one type of aircraft, point-to-point short-haul travel, no in-flight meals, rapid turnaround time, very high aircraft utilisation, direct sales, cost-conscious customer segments and extensive sub-contracting. easyJet’s managers identified three of its
nearest low-cost competitors and the strategy of each of these airlines is detailed in
the case study. But easyJet also experienced direct retaliation from large flag carri-
ers like KLM and British Airways (Go). These challenges faced easyJet’s owner,
Stelios Haji-ioannou, as he signed a $500 m contract with Boeing in July 1997 to
purchase 12 brand new 737s.

Imagine yourself now in Mr. Haji-ioannou’s role. Is it really going to be feasible
to fill those expensive new planes? In his mind is a bright new business idea, a
creative new segmentation of the air travel market to be achieved through cost leader-
ship and aimed at customers who are interested in “jeans not business routines”.
Feasibility checks of strategy are natural territory for business simulators, especially
dynamic, time-dependent, strategy problems such as rapid growth in a competitive
industry. At the time there were differences of opinion within the industry and even
among easyJet’s management team. Some industry experts had a dismal view of
easyJet’s prospects (in stark contrast to the founder’s optimism), dismissing the fledgling
airline with statements such as “Europe is not ready for the peanut flight”.

To bring modelling and simulation into this debate we have to visualise the
dynamic tasks that face Mr. Haji-ioannou and his team in creating customer aware-
ness (How do you attract enough fliers to fill 12 planes?), and dealing with retalia-
tion by rivals (What if British Airways or KLM engage in a price war, could they
sustain such a war, what would provoke such a response?). The starting point is a
map of the business, a picture created with the management team, to think with
some precision about the task of attracting and retaining passengers and the factors
that might drive competitor retaliation.

2.2.5.2 Winning Customers in a New Segment: A Process that Involves
Stock Accumulation and a Reinforcing Feedback Loop

Recall that the building blocks of system dynamics models are stock accumula-
tions, causal links and feedback loops. Causal links show simple cause and effect
relationships. Feedback loops depict closed paths of cause and effect and are of
special importance because they generate dynamics. Feedback loops can be either
reinforcing or balancing. Reinforcing loops are responsible for growth dynamics
whereas balancing loops are responsible for goal-seeking dynamics and oscilla-
tions. By combining stock accumulations, causal links and feedback loops it is
possible to create visual models of a wide variety of dynamic strategic business
situations, including easyJet’s $500 million gamble.

Figure 2.26 uses one stock accumulation, one reinforcing feedback loop and several
causal links to show how a start-up airline attracts new passengers and communicates
its new low-cost, no-frills service to the flying public. The marketing task is
far-from-trivial, because when you think about it (and modelling really forces you to
think hard about the practical details that underpin strategy) the company has to spread
the word to millions of people if it is to fill 12 brand new 737s day after day.

Potential passengers are shown as an asset stock representing the cumulative
number of fliers who have formed a favourable impression of the start-up airline.
Note that these passengers have not necessarily flown with easyJet, but would if they could. This rather abstract way of thinking about passengers is a convenient simplifying assumption that enables us to focus on growth of interest in low-cost flights without the need to model the detailed operations of the company. Bear in mind however that the scope of a model always depends on its purpose. For example a model to study the growth of the whole airline (rather than simply growth of potential passengers) would include the company’s internal operations such as hiring and training of staff and investment in planes, as in Sterman’s (1988) well-known People Express Management Flight Simulator.

The number of potential passengers starts very small (just 5,000 in the model) and grows over time. But how does growth take place? The remaining parts of the figure show the factors that determine both the increase and loss of passengers. In practice this information comes from the management team, coaxed-out by a facilitator who is helping the team to visualise the business.

The driver of growth is a reinforcing feedback loop shown at the centre of Fig. 2.26 and labelled ‘Growth Engine’. In this loop potential passengers attract new converts through positive word-of-mouth. The more potential passengers, the

---

Potential Passengers\( (t) = \) Potential Passengers\( (t-dt) + (\text{Increase of Potential Passengers} - \text{Loss of Potential Passengers}) \times dt \)

INIT Potential Passengers = 5000

Increase of Potential Passengers = Potential Passengers \times Conversion Ratio + Marketing Spend \times 1000 \times Marketing Effectiveness

Conversion Ratio = GRAPH(Relative Fare)

Marketing Spend = 2500[£thousands per year]

Marketing Effectiveness = 0.05  {Passengers wooed per £spent}

Fig. 2.26  Creating awareness of low-cost flights among potential passengers: word-of-mouth and marketing
greater the rate of increase of potential passengers. The increase of potential passengers then accumulates in the stock of potential passengers leading to even more potential passengers and a greater rate of increase in potential passengers, thereby completing the reinforcing loop. The strength of word-of-mouth is captured in a concept called the conversion ratio, which itself depends on relative fare. As relative fare increases the conversion rate decreases, a causal link with a ‘−’ sign on the arrow head to indicate negative polarity.

These effects are captured algebraically in the equation for increase of potential passengers. The first part of the equation states that the increase of potential passengers depends on the product of potential passengers and the conversion ratio. Intuitively the lower easyJet’s fare relative to established rivals the higher the conversion ratio and the more potent is word-of-mouth. An exceptionally low fare is a talking point among the travelling public, just as happened in real life.8 Such a relationship would normally be sketched as a graph, based on expert opinion from the management team. The shape of the graph can be seen by browsing the Fliers Mini-Sim on the website at http://www.iseesystems.com/community/downloads/OpenUniversity.aspx. The graph shows that when easyJet’s fare is just 30% of rivals’ fare the conversion ratio is 2.5, meaning that each potential passenger converts 2.5 new potential passengers per year. However at 50% of rivals’ fare the conversion ratio is reduced to 1.5 and at 70% it is only 0.3. Eventually, if easyJet’s fare were to equal rivals’ then the conversion ratio would be zero because a standard fare cannot sustain word-of-mouth.

The increase of potential passengers is also influenced by marketing spend, another causal link. This link is formulated as the product of marketing spend and marketing effectiveness (shown in the second part of the equation for increase of potential passengers). Marketing spend is set at a default value of £2.5 million per year. Marketing effectiveness represents the number of new potential passengers per marketing £ spent. It is set at 0.05 passengers per £, so marketing brings 125,000 potential passengers per year (2.5 million per year * 0.05).

The loss of potential passengers depends on service reputation. The lower service reputation, the greater the churn. The greater the churn the more the loss of passengers. Industry specialists say that service reputation depends on ease-of-booking, punctuality, safety, on-board service, and quality of meals. For short-haul flights punctuality is often the dominant factor. The model does not represent all these factors explicitly but simply represents service reputation as a stock accumulation that can be initialised anywhere on a scale between 0.5 (very poor) and 1.5 (very good). If reputation is very good then fliers retain a favourable impression of the airline, so the annual loss of potential passengers is small, just 2.5% per year – an assumption made in the graph function for the churn. If reputation is poor then the loss of potential passengers per year is damagingly high, up to 100% per year. Notice there is no inflow or outflow to reputation even though it is a stock variable. The reason is that the factors driving change in reputation are outside the boundary of the model.

---

8In some cases very low fares may deter passengers due to concerns about safety. But in this particular case easyJet was flying a fleet of brand new 737s which instilled confidence.
2.2.5.3 Retaliation by High-Cost Rivals: A Process That Involves Stock Accumulation and a Balancing Feedback Loop

Figure 2.27 shows one possible way to visualise the retaliatory response of powerful European flag carriers to low cost airlines in the early years. It is important to emphasise here the phrase one possible way, because there are many ways that a management team such as easyJet’s might think about competitors. Part of the team model-building task is to achieve the simplest possible shared representation, drawing on the sophisticated (and sometimes conflicting) knowledge of the team members. A fundamental question is whether it is necessary to model competing firms in-depth. Do you really need a detailed portrayal of British Airways or KLM to understand the threat such rivals might pose to the feasibility of easyJet’s growth strategy?

The leader of a team-modelling project should not impose a rigid answer on this question of how much detail to include. The modeller should be sensitive to the opinions of the management team while always striving for parsimony. After all to achieve buy-in the model must capture managers’ understanding of their world in their own vocabulary. In these situations it is useful to bear in mind that experienced business leaders themselves simplify their complex world. If they did not then it would be impossible to communicate their plans. Good business modelling, like good business communication, is the art of leaving things out – focussing only on those features of reality most pertinent to the problem at hand.

\[
\text{Rivals Fare}(t) = \text{Rivals Fare}(t - \Delta t) + (\text{Change in Rivals Fare}) \times \Delta t
\]

\[
\text{INIT Rivals Fare} = 0.25 \ \{\text{£/passenger mile}\}
\]

\[
\text{Change in Rivals Fare} = (\text{Fare Set by Startup} - \text{Rivals Fare}) / \text{Time to Change Costs}
\]

\[
\text{Fare Set by Startup} = 0.09 \ \{\text{£/passenger mile}\}
\]

\[
\text{Time to Change Costs} = 4 \ \{\text{years}\}
\]

\[
\text{Relative Fare} = \text{Fare Set by Startup} / \text{Rivals Fare} \ \{\text{dimensionless}\}
\]

**Fig. 2.27** Rivals and relative fare
Figure 2.27 shows just enough about competitors to indicate how, collectively, they could stall easyJet’s growth ambitions. Recall that word-of-mouth feedback relies for its contagion on the start-up’s fare being much lower than rivals. But what if competing firms try to match the start-up’s low price? The figure shows how such price equalisation might take place. At the heart of the formulation is a balancing loop labelled ‘Restructuring’. Rivals’ fare is shown as a stock that accumulates the change in rivals’ fare which in turn depends on three factors: the fare set by the startup, rivals’ fare and the time to change costs, all depicted as causal links. The use of a stock accumulation implies that it takes time and effort for the established airlines to lower their fares. They cannot reduce fares until they cut costs, and a flag carrier like BA may take years to achieve cost parity with a low-cost start-up. The process of achieving cost parity is essentially a goal-seeking process represented by the balancing loop.

To understand the operation of the balancing loop let us suppose, for the sake of argument, that rivals begin with an average fare of 25 pence (£0.25) per passenger mile and set themselves a goal for average fare of only nine pence (£0.09) per passenger mile – equal to the average fare set by the start-up. (Of course nowadays all airlines use revenue management systems with variable fares. But our focus is on the huge discounts originally offered by low-cost airlines that were available on most seats and enabled easyJet to grow. So a very low fixed fare for the start-up is a reasonable simplifying assumption.) The magnitude of the underlying cost equalisation task is now clear – it is the 64% difference between rivals’ initial fare of 25 pence (£0.25) and easyJet’s fare of nine pence. Such an enormous change can only be achieved through major restructuring of the business. The change in rivals’ fare is controlled by the ‘restructuring’ balancing loop that gradually reduces the fare to equal the fare set by the start-up. The pace of restructuring depends on the time to change costs. Normally one would expect this adjustment time to be several years, and in the model it is set at 4 years. The equations show a typical asset stock adjustment formulation. The change in fare is equal to the difference between the start-up’s fare and rivals’ fare divided by the time to change costs. This expression takes a negative value as long as rivals’ fare exceeds the start-up’s fare, thereby leading to a fare reduction. So, at the start of the simulation, the change in fare is

---

9 Rivals are portrayed at a high level of aggregation. The purpose is to capture in broad (but dynamically accurate) terms how rival airlines respond to price competition.

10 Large carriers will match low seat prices regardless of cost by providing some seats at a discount. Price cuts can be implemented very quickly through on-line yield management systems that allow dynamic pricing according to load factors. But narrowly targeted discounts are an ineffective weapon for companies like BA and KLM in the competitive fight with low-cost airlines. For example, out of 150 seats there may be only 15 cheap ones. For very popular flights there are no cheap seats at all. Only cost parity can deliver competitive prices and profitability in the long-term for large carriers catering to a growing population of price-conscious fliers.

11 An empirical study of cost and productivity convergence among US airlines, conducted by Peter Belobaba from MIT’s International Centre for Air Transportation, confirms significant cost convergence between Network Legacy Carriers and Low Cost Carriers spread over several years.
(0.09–0.25)/4 which is a brisk reduction rate of £0.04 per passenger mile per year. This rate prevails over the first computation interval to arrive at a new and lower fare for the next computation interval, and so on as the simulation proceeds.

2.2.5.4 Feedback Loops in the easyJet Model

Figure 2.28 summarises the main feedback loops in the model, including the two loops described above and two more loops that capture route saturation and churn. In the centre of the figure is the reinforcing growth engine from Fig. 2.26. More potential passengers lead to more conversion from word of mouth, a greater increase of potential passengers, more potential passengers, and so on.

In the bottom right of the figure is the important balancing loop from Fig. 2.27, involving restructuring of costs, which determines rivals’ fare. There are just two concepts in the loop: rivals’ fare and cost cutting rate. The dynamic significance of the balancing loop is that it tends to equalise rivals’ fare with the start-up’s fare. As a result relative fare (defined as the ratio of start-up’s fare to rivals’ fare) converges gradually to parity thereby reducing the strength of word-of-mouth in the reinforcing loop.

These two loops form the core of the model and are central to the evaluation of easyJet’s start-up strategy. Qualitatively, if the reinforcing loop is strong (and stimulates rapid growth) while the balancing loop is weak (and leads to very slow price equalisation) then easyJet’s $500 million gamble is likely to succeed and the company will fill its planes. However, if the balancing loop is strong (and price

Fig. 2.28 Feedback loops for the launch of a low-cost airline
equalisation happens quickly) then the window of opportunity for rapid growth is much reduced and easyJet’s gamble may fail.

In addition to the two loops depicted in bold there are two further loops in Fig. 2.28 that capture the effects of route saturation and churn on passenger interest. These extra loops are peripheral to the immediate question of whether or not easyJet can fill 12 planes, but are important in the long run to ensure realistic limits to the growth of potential passengers in the region served by the fledgling airline. At the top of the figure is a balancing loop (labelled ‘Limiting Process’) in which route saturation eventually restricts the increase of potential passengers. Finally in the centre-right is a balancing loop (labelled ‘Churn’) showing the effect of the start-up’s service reputation on the loss rate of potential passengers.

Of course this brief model of passengers and fares is a sketch of a more complex reality. Nevertheless, it contains sufficient detail for an informative team discussion about passenger growth and price retaliation. And when simulated the model contains sufficient dynamic complexity to yield thought-provoking growth scenarios that help management to rehearse strategy.

### 2.2.5.5 Strategy and Simulation of Growth Scenarios

The purpose of the model is to investigate easyJet’s $500 million gamble to purchase twelve brand new Boeing 737s. Is it wise to order so many planes? Will it be possible to fill them? And assuming a large potential market for low-cost air travel, will easyJet be able to capture a big enough slice? A rough calculation suggests the airline needs one million fliers if it is to operate 12 fully-loaded aircraft\(^{12}\) – which is a lot of people. What combination of word-of-mouth and marketing will attract this number of potential passengers? How long will it take? What are the risks of price retaliation by rivals? These are good questions to explore using the what-if capability of simulation.

Figure 2.29 shows simulations of the growth of potential passengers over the period 1996–2000 under two different approaches to marketing spend (bold and cautious) and under the assumption of slow retaliation by rivals. Bold marketing spend is assumed to be five times greater than cautious spend (at £2.5 million per year versus £0.5 million per year). In both cases the horizontal straight line shows the ‘required’ number of passengers to fill 12 planes. This line is a useful reference

---

\(^{12}\)Let’s assume each aircraft carries 150 passengers and makes three round-trip flights a day. So a fully loaded plane needs 900 passengers each day \((150\times3\times2)\). A fully loaded fleet of 12 planes needs 10,800 passengers a day, or 3,888,000 passengers each year, which is very nearly four million. If we make the further assumption that each potential passenger is likely to fly the available routes twice a year on round-trip flights, then the start-up airline needs to attract a pool of almost one million fliers to ensure commercially viable load factors. This rough calculation is typical of the sort of judgmental numerical data required to populate an algebraic model. Perfect accuracy is not essential and often not possible. The best estimates of informed people, specified to order-of-magnitude accuracy (or better), are adequate, drawing on the informal but powerful knowledge base derived from experience.
against which to compare the number of potential passengers. If and when potential passengers exceed required passengers, the strategy is deemed feasible.

Consider first the timeline for bold marketing in the top half of the figure. The simulation begins in 1996 with a very small number of potential passengers – just 5,000. The fledgling airline is virtually unknown to the flying public, despite its ambitions. In the first year of operation, bold marketing brings the airline to the attention of a growing number of fliers. By the end of 1996 there is a band of several hundred thousand enthusiastic supporters. Moreover, this band of supporters is beginning to recruit more followers through positive word-of-mouth. In the interval 1997–1998 the number of potential passengers rises sharply as

---

**Fig. 2.29** Simulations comparing bold marketing (*top chart*) with cautious marketing (*bottom chart*). Assuming slow retaliation
word-of-mouth continues to stoke exponential growth. By mid-1997 the number of potential passengers has reached the target of one million required to fill the fleet. In the remainder of the year, reinforcing growth continues. There is a huge leap of more than one million potential passengers in the last 6 months of 1997 as the powerful engine of growth continues to gather momentum. Then, in the second quarter of 1998, growth ceases abruptly as the airline’s message reaches all 3.5 million fliers in the imagined catchment region it serves.

The strategically important part of the timeline is the growth phase between the start of 1996 and early 1998. Bold marketing coupled with strong word-of-mouth unleashes a powerful engine of growth which, in classic exponential fashion, begins small (and therefore invisible) and snowballs rapidly after 18 months.

Now consider the timeline in the bottom half of Fig. 2.29, which traces the build-up of potential passengers from cautious marketing. Spend is cut by four-fifths from £2.5 million a year to only £0.5 million a year. As before, the simulation starts in 1996 with only 5,000 potential passengers. In the first year the airline wins few passengers – not surprising because marketing spend is much reduced. In the second year there is healthy growth in passengers, despite the low marketing spend. Word-of-mouth is now beginning to draw-in lots of new passengers. Once the growth engine is primed it gets rolling and in the second quarter of 1998 carries the airline’s passenger base beyond the target required to fill the fleet. Growth continues into 1999 until nearly all 3.5 million fliers are aware of the new low-cost service. Cautious marketing simply defers growth (by comparison with bold marketing) but doesn’t seem to radically alter the ultimate size of the passenger base. One can begin to appreciate a persuasive rationale for caution. By the year 2000 the simulated airline has saved £8 million in marketing spend (4 years at an annual saving of £2 million) yet has still got its message out to 3.5 million fliers!

Figure 2.30 shows the same two marketing approaches (bold and cautious) under the assumption that rivals retaliate quickly. Price equalisation happens in half the time previously assumed and as a result both timelines are noticeably changed by comparison with the base case. But from the viewpoint of strategic feasibility the bold marketing timeline tells much the same story as before. At the start of 1996 the airline is almost unknown among the flying public, and by the third quarter of 1997 it has attracted enough potential passengers to fill 12 planes. Fast-acting rivals seem unable to prevent this rise of a new entrant from obscurity to commercial viability, though price equalisation measures do curtail the ultimate dissemination of the start-up airline’s low-price message.

A strategically significant change is observable in the timeline for cautious marketing. The startup airline is no longer able to fill its planes because it is unable to attract passengers. The rise from obscurity to prominence never happens. Cautious marketing attracts few converts and fails to ignite word-of-mouth. By the time the low-price message has reached a few hundred thousand fliers (at the end of 1997) it is no longer distinctive. Rivals are low price too. If this future were easyJet’s its planes would be flying half-empty and it would be losing money. Fast retaliation can prove fatal in a word-of-mouth market.
2.2.5.6 Using the Fliers Simulator to Create Your Own Scenarios

The Fliers simulator enables you to explore a variety of scenarios for a start-up low cost airline. You can replay the simulations shown above, create new scenarios, and investigate the behaviour of many more variables. Open the model called Fliers Mini-Sim on the website at http://www.iseesystems.com/community/downloads/OpenUniversity.aspx to see the opening screen as shown in Fig. 2.31. There is a time chart for potential passengers and required passengers, numeric displays for potential passengers and relative fare, and slide bars for marketing spend and time to change costs. Marketing spend is 2,500 (in £ thousands per year) and the time to change costs is 2 years. The top chart shows bold marketing of £2.5 million per year, and the bottom chart shows cautious marketing of £0.5 million per year.

Fig. 2.30 Simulations comparing bold marketing (top chart) with cautious marketing (bottom chart). Assuming fast retaliation.
change costs is 4 years. These are the conditions for the base case scenario of bold marketing and slow retaliation already seen in Fig. 2.29.

To get started press the ‘Run’ button without altering either of the slide bars. The first year of the simulation plays out. Scroll through the time charts to view the behaviour of the conversion ratio, the effect of route saturation, churn and the increase/loss of potential passengers. Press the Run button again to see the next simulated year and so on to the end of the simulation in the year 2000. For a guided tour of the simulation press the scenarios button on the left. A new screen appears containing a menu of pre-prepared scenarios. Press the large green button for a year-by-year analysis of the base case. At the end of the analysis press ‘scenario explorer’ to return to the opening screen. Then conduct your own experiments with other combinations of marketing spend and time to change costs. At any time you can learn more about the simulator by pressing the navigation buttons on the left of the screen. The introduction is a review of the easyJet case and the feedback structure of the model. The scenarios button offers a guided tour of the four pre-prepared scenarios already covered in Figs. 2.29 and 2.30. ‘Browse model’ allows you to see the detailed model structure and documented equation formulations.

2.2.6 Excerpts from ‘Orchestras in a Complex World’

Sometimes casual loop diagrams are used in a purely qualitative way, without algebraic modelling and simulation. Although the emphasis of this chapter has
been on mapping that leads to simulation, it is also useful for readers to see that conceptual maps are helpful in their own right: for expanding the boundary of people’s thinking about organisations; and for providing an overview from which novel insights may arise.

To illustrate I present an application to orchestra management carried out by Bernhard Kerres while he was an associate with Booz-Allen Hamilton in Munich. Over a period of 3 years he worked with various orchestras in Europe to assist in the development of their strategic agendas. In the course of the study he spoke with orchestra managers, concert promoters, musicians, agents and others close to the industry. Together they explored the questions of what is success for an orchestra and how can an orchestra become successful. Drawing on his experience as a professional musician and knowledge of system dynamics from an MBA at London Business School he was in a good position to help orchestra managers and other stakeholders to address these questions.\textsuperscript{13} Collectively the interviewees identified five major indicators for successful orchestras:

\begin{itemize}
  \item High quality orchestral concert performances with the ability to attract and retain excellent orchestra musicians, as well as guest artists and conductors
  \item Challenging and interesting programming which attracts audiences and raises the interest of new audiences
  \item Attracting well-qualified managers and staff, and also enthusiastic volunteers and supportive sponsors
  \item Maintaining a media profile, including recordings and broadcasts, as well as favourable reviews
  \item Successful outreach and education work through provision of musical services to their communities, with an outcome of raising the understanding and appreciation of music
\end{itemize}

Causal loop diagramming was used to show how these indicators are related to each other. Here I present an \textit{edited subset} of the diagrams that appeared in Bernhard Kerres’ published article. Note how he communicates complexity. He adds interlocking loops one-by-one and writes a vivid accompanying narrative that is well-grounded in the real-world situation.

\subsection*{2.2.6.1 Success of Performances and Quality of Orchestra}

To build a conceptual map for orchestras in larger cities, the success of metropolitan performances seems to be a good starting point. What is a successful performance? What makes it successful? Successful performances can be seen mainly in two ways: artistic success and financial success.

\textsuperscript{13}I am grateful to Bernhard Kerres (1999) for this example which is based on excerpts from an article entitled ‘Orchestras in a complex world’ first published in \textit{Harmony}, 8, pp. 45–58 (Forum of the Symphony Orchestra Institute). Bernhard is now Intendant and CEO of the Vienna Konzerthaus, one of the most active concert houses in the world.
Artistic success includes the quality of the performance and the challenge of the programme. Financial success includes the ability to sell tickets for the performance and to attract sponsorship. The comparable dimension of intermediate or longer-term artistic and financial success would be the ability to sell season subscriptions and to increase the audience base.

When considering what makes the actual success, audience attraction is a major point as shown in Fig. 2.32. Financial and artistic success can only be achieved if audiences are attracted. Audiences are often attracted by their interest in the programme and/or the attraction artists hold, including the fundamental quality of the orchestra. Programmes and artists are often cited as the main reasons why audiences attend concerts. Sometimes the venue itself plays a role. The state-of-the-art Benaroya Hall in Seattle, Symphony Center in Chicago, or the Konzerthaus in Vienna are attractive and unique places in themselves, and attract audiences who want to say “I’ve been there”.

And one should remember that audiences are not the only people who come to listen to a concert, broadcast, or recording. Audiences also include supporters, volunteers, and others who endorse the work of an orchestra. This wider definition of audience is critical for the success of orchestral institutions. Without the support from the wider audience, an orchestra would be limited to silent listeners. It would be hard for an orchestral institution to become a lively organisation which attracts great artists and musicians, or to be successful in the longer term.

And undoubtedly the media have a role in attracting audiences for performances. But what exactly is that role? Media includes print, as well as recordings, broadcasts and many other forms. Media is a very large industry in itself and often crosses paths with the music industry. Media attention includes not only reviews, but also any form of publicised information about the orchestral organisation’s activities. This can range from advertising at the local bus station to dedicated slots on the local radio station.

Technical developments in the media industry over recent years have lowered barriers to the media world, but also raised the level of competition. Orchestral institutions face the challenge of how to use these developments to their advantage.
The options are immense, and might include selling recordings over the internet or entering into partnerships with various media companies. Just consider the Berlin Philharmonic which has recently started to make all their concerts available on the internet with great sound and video quality in a subscription model.

With these thoughts in mind we have now covered the first part of the conceptual map shown in Fig. 2.32. The map establishes connections between the quality of an orchestra and the success of performances. However it does not adequately explain the influence of the media on success (shown as a dotted line with an accompanying question mark). Neither does the figure yet show any reinforcing feedback loop. If one or more such loops can be established for an orchestra, that orchestra would have found a success engine to drive growth.

2.2.6.2 The Importance of Brand

So far we have not spoken about the “brand” of the orchestra. In today’s world a brand for an orchestra is just as important as for any other good. Such examples as Virgin demonstrate how powerful brands can be. But there are also examples of powerful brand names in the orchestra world. Such orchestras as the Vienna Philharmonic or the Berlin Philharmonic are associated with world-class quality and other attributes. The names of these orchestras have developed into brand names, even if these orchestras do not actively promote their brands. And so have certain artists like Anna Netrebko, Lang Lang and many others.

But what lies behind a brand name? A brand relies on the image it generates in people’s minds. A brand links the values of a product or organisation with the qualities people associate with the product or the organisation. We therefore should consider not only such well-known brands as Coca-Cola. The local shop in a small town actually has a brand because the local population links the image of the shop with the values the shopkeeper represents. The only differences are that fewer people know the brand, and it may not be as well managed as Coca-Cola.

One example of an orchestra developing its image into a brand is the Detroit Symphony. With its surprise encores and the friendliness it exhibits towards its audiences, the Detroit Symphony is creating a certain favourable image in the minds of its audiences. It is building a brand with this image to differentiate itself from other orchestras, and from other performing arts groups in Detroit. The correct conclusion is that orchestral institutions in any city have to think hard about the qualities and values they want people to think of when they hear or see the orchestra.

Take the example of the Florida Orchestra, which works hard on its image of being informal and creative. On one occasion the orchestra performed an all-Frank Zappa concert as part of its frequent testing of the boundaries among classical music, jazz and pop music. The Washington Post reported that. “Roars of applause followed every piece…. Symphony patrons in tuxedos edged past colourful eccentrics decked out in Willie Nelson braids and Harley leathers”.

We now add brand to the conceptual map as shown in Fig. 2.33. Because brand recognition can be measured, it is a good parameter for a conceptual map. Brand establishes the first feedback loop in our model: the stronger the brand recognition of an orchestra, the higher the attraction to audiences and the better the success of performances. This feedback loop is represented by the symbol R1 on our map and is labelled ‘Brand Growth Engine’ to indicate its potential to generate growth. However it is important to note that the same loop could change from a virtuous circle to a vicious circle under adverse circumstances. For example if the success of performances is low, then brand recognition could decline and even become negative. Negative brand recognition can lead to lower audience attraction, which in turn can lead to less successful performance. Keep in mind this possible switch in behaviour as we extend the map by adding more reinforcing loops.

Figure 2.33 also helps to better understand the media’s role in the success of performances. Media – in its full variety – directly influences brand recognition and indirectly affects success through loop R1. The more an orchestral institution appears in articles, broadcasts, shows and reviews, the higher the brand recognition and the greater the knock-on consequences to success.

2.2.6.3 Attracting Musicians

So far the map is missing one element vital for an orchestra’s success. An orchestra could not exist without its musicians. Figure 2.34 shows how musicians can be attracted. Musicians consider important the orchestra for which they play. They take into consideration the brand of the orchestra as well as the soloists and conductors with whom they work.

If an orchestra is attractive to well-known conductors, it will also be attractive to musicians. Attracting good and enthusiastic musicians is critical for the success of
an orchestra and its performances and establishes a second reinforcing feedback loop R2, labelled ‘Best Musicians’.

2.2.6.4 Success with Fundraising

Our map so far has not touched upon a very important issue for any arts organisation: fundraising. In general, few orchestras in the United States ever really experienced the system of public funding which was well known until recently in Europe and Canada. In Europe, funding for the arts was historically reserved for the sovereign. The shortfalls in state households and the focus on other issues have led to a steady decrease in public funding in most European countries. Private fundraising has now become as important in Europe and Canada as it has been historically in the United States. A conceptual map for orchestra organisations needs to take this development into account.

Fundraising success – from individual, corporate and public sources – depends heavily on the brand recognition of the arts organisation. An organisation with a good brand recognition will also be able to attract the right supporters and volunteers to make fundraising a success. The success of the orchestral institution itself depends on the ability to raise sufficient funds. Fundraising success must therefore be included in the conceptual map.

As Fig. 2.35 shows, fundraising success depends heavily on brand recognition. It is easier to raise funds for an organisation which is well known and well thought of than for an unknown organisation. Well-known orchestras can attract higher levels of funding and can also attract prominent individuals to leadership of their fundraising campaigns.

Survival for lesser-known organisations is a real issue, especially in Europe. Lesser-known organisations in countries which traditionally had high public funding face not only drastic reductions in public funding, but also see corporate sponsors attracted to the top institutions (which, ironically, still receive a certain level of
public funding). Public and private funding become focussed on a few well-known organisations, leaving fewer funds available for lesser-known institutions.

Success in fundraising starts another reinforcing feedback loop R3. Only if enough funds are available will well-known soloists, conductors and musicians be attracted to perform with the orchestra. High-level artistry is necessary to develop audiences and to generate sufficient media interest. This process again leads to better brand recognition.

Similar to brand recognition, fundraising seems to be a key success factor for orchestras in today’s environment. Many orchestral organisations in the United States have professional fundraising staffs, either in-house or outsourced. Orchestra organisation board members in the United States take active roles in fundraising, often giving significant donations to their organisations.

Private donors are the strongest supporters of US arts organisations. They may not necessarily be interested in a well-marketed brand, but they are interested in the image behind the brand. If they see their own interests and values represented in the image behind the brand, they will be inclined to support a particular orchestral institution.

The climate in Europe is very different. Fundraising is rather new. Some organisations in the United Kingdom are taking the lead. Nevertheless, many boards, if they exist at all, see their roles primarily in governance and not in fundraising. A learning process will obviously be necessary.

### 2.2.6.5 Conclusions from Orchestra Study

A conceptual map represents ways in which the organisation’s main features and activities interrelate with one another, and with the environment in which the organisation functions. Building a conceptual map is normally done in an iterative team effort. The map represents the group’s consensus of the operating environment.
A successful conceptual map requires the support of the whole team. Therefore, it is valuable to work not only with the management and the board of an orchestral organisation, but also to include supporters, sponsors and representatives of audiences. The success of an orchestra is not based on a few people on the orchestra’s payroll. In today’s world, staff, musicians, volunteers, audiences, supporters, and many others take an active interest in the future of their orchestral institutions. Incorporating their views, with the help of a trained facilitator and map builder, increases the chances for a successful process.

2.3 Summary and Conclusion: An Overview of the Modelling Process

The previous examples have covered the main concepts and tools used in system dynamics. In summary, five steps of modelling can be identified as shown in Fig. 2.36. Usually there is lots of to-and-fro between the steps as understanding of the situation improves by sketching diagrams, quantifying concepts, writing friendly algebra, and making simulations. Step 1 is problem articulation. It is the most important step of all because it shapes the entire study. Here the modeller or modelling team identify the issue of concern, the time frame, the level of analysis (business unit, firm, industry, etc.), the boundary of the study and the likely scope of factors involved. Step 2 is a dynamic hypothesis, a preliminary sketch by the modeller of the main interactions and feedback loops that could explain observed or anticipated performance. Step 3 is formulation, the transformation of a dynamic hypothesis into a reasonably detailed diagram of feedback processes and corresponding algebraic equations. Step 4 is testing. The model is simulated to see whether or not its behaviour over time is plausible and consistent with available

![Fig. 2.36](image-url) Modelling is an iterative learning process (Sterman 2000). Business Dynamics: Systems Thinking and Modeling for a Complex World, McGraw Hill, with permission from The McGraw-Hill Companies
evidence from the real world. Step 4 fixes errors and begins to build confidence in the model’s integrity. Step 5 is policy formulation and evaluation. By now there is confidence that the model’s structure is sound and that it is capable of reproducing the dynamic symptoms of the original problem. So attention shifts to policy changes intended to improve performance and to alleviate the perceived problem. The new policies are then simulated to see how well they work.

Notice these steps are shown as a cycle and not as a linear sequence. The web-like symbol in the middle of the diagram and the circle of arrows around the edge mean that iteration is a natural and important part of the process. For example it is common for modellers to revise the problem and model boundary as they develop a dynamic hypothesis and causal loops. So step 2 influences step 1. Similarly formulation and testing can reveal the need for new equations or new structure because simulations contradict common sense or else reveal that the original dynamic hypothesis is incapable of generating observed or expected behaviour over time. So steps 3 and 4 can influence steps 1 and 2 or each other.

Consider such iterations as they arise in the drug-related crime model. According to the dynamic hypothesis, escalating crime is attributable to police drug busting that removes drugs (and drug dealers) from the streets. A side-effect is to push up the street price of drugs and this price inflation inadvertently forces addicts to commit more crime, leading to more drug busting and so on. The structure is a reinforcing loop and the resulting simulator (based on formulations outlined in Figs. 2.18–2.25) shows that crime escalation is possible given reasonable operating assumptions about the police department, street market, the community and addicts themselves. However, when the simulator is run for 5 or more years this logic is pushed beyond the limits of common sense and reveals a world in which the price of drugs is sky high, crime has increased sixfold and the supply of drugs on the street is negative!

A modeller faced with these contradictions returns to the model’s assumptions to find the fallacy. A few possibilities come to mind. The simplest, and least disruptive to the integrity of the model, is that police effectiveness in drug busting is not constant (as assumed) but depends on the supply of drugs on the street. As supply is reduced through drug seizures it becomes more and more difficult for police to trace the few drugs that remain – an example of the ‘law of diminishing returns’. This formulation requires a new causal link and a graphical converter that shows police effectiveness as a non-linear function of the supply of drugs on the street. Another more radical idea is to include the dynamics of supply. The current model assumes the total supply of drugs is fixed, so drug seizures create a permanent shortage on the street. But if the street price is high then, sooner or later, the supply of drugs will increase to compensate for drug busting, thereby re-establishing an equilibrium of supply and demand. In other words the dynamic hypothesis needs to be modified and the boundary of the model expanded in order to create plausible long-term dynamics.

---

14 The simulations are not included in this chapter. Readers who wish to see them should refer to Chapter 3 of Strategic Modelling and Business Dynamics.
2.3.1 Dynamic Hypothesis and Fundamental Modes of Dynamic Behaviour

From a modeller’s perspective a dynamic hypothesis is a particularly important step of ‘complexity reduction’ – making sense of a messy situation in the real world. A feedback systems thinker has in mind a number of structure-behaviour pairs that give valuable clues or patterns to look for when explaining puzzling dynamics. Figure 2.37 shows six fundamental modes of dynamic behaviour and the feedback structures that generate them.

The trajectories in the top half of the diagram arise from simple feedback processes. On the left is pure exponential growth caused by a single reinforcing feedback loop in isolation. In the centre is pure goal seeking behaviour caused by a balancing loop. On the right is s-shaped growth that occurs when exponential growth hits a limit. In this case a reinforcing loop dominates behaviour to begin with, and then later (due to changing conditions) a balancing loop becomes more and more influential.

The trajectories in the bottom half of the diagram arise from more complex feedback processes. On the left is classic oscillatory, goal-seeking behaviour with repeated overshoot and undershoot of a target, caused by a balancing loop with a time delay. In the centre is growth with overshoot, a pattern of behaviour where growth from a reinforcing loop hits a limit that is not immediately recognised. This lagged limiting effect is represented as a balancing loop with delay. On the right is overshoot and collapse, which is a variation on growth with overshoot. But here the limit itself is a floating goal that adds an extra reinforcing loop. This set of six structure-behaviour pairs is not exhaustive but illustrates the principle that any pattern of behaviour over time can be reduced to the interaction of balancing and reinforcing loops.

Some of the most intriguing and complex dynamics arise in situations where multiple feedback loops interact and each loop contains time delays and non-linearities. Even quite simple models with two or three interacting loops can prove to be very

![Fig. 2.37 Dynamic hypothesis and fundamental modes of dynamic behavior (Sterman 2000). Business Dynamics: Systems Thinking and Modeling for a Complex World, McGraw Hill, with permission from The McGraw-Hill Companies](image)
interesting as we saw in the airline model presented earlier. The main point for now is to realise that all such models take shape in a structured yet creative process of discovering feedback processes in everyday affairs.

### 2.3.2 A Spectrum of Model Fidelity

Models range in size from large-and-detailed to elegantly small and metaphorical. The spectrum is illustrated in Fig. 2.38. On the left-hand side are realistic high-fidelity simulators epitomised by aircraft flight simulators used to train pilots. They are realistic enough for pilots to practice take-offs and landings and to prepare for emergencies such as engine failure. Often people expect business and public policy models to be similarly realistic; the more realistic the better. But very often small and simplified models are extremely useful as metaphors for more complex situations.

My favourite example of a metaphorical model is a simulator of Romeo and Juliet intended for high school students studying Shakespeare in English literature classes. Clearly a simulator cannot possibly replicate Shakespeare’s play, but it can encourage students to study the play more closely than they otherwise would. By simulating the waxing and waning of love between Romeo and Juliet, students become curious about romantic relationships, both in the model and the play. A metaphorical model is small and can be explained quickly. The Romeo and Juliet simulator fits on a single page and involves just a handful of concepts, a far cry from the large and detailed model that lies behind an aircraft flight simulator. It is important to realise that business and public policy models typically lie somewhere in the middle of this spectrum of model fidelity, as indicated by the oval in Fig. 2.38.

![Fig. 2.38  Modelling and realism – a spectrum of model fidelity](image)

15The Romeo and Juliet simulator is described in chapter 6 of an edited book entitled *Tracing Connections, Voices of Systems Thinkers* (Morecroft 2010). See the list of references for a full citation.
2.3.3 Growth Strategy in Low-Cost Airlines: A Small Model and a Much Larger One

The easyJet case gives us a taste of a small but nevertheless quite insightful model – ‘a back of the envelope model’ – to address a dynamic challenge in a rapidly evolving market where timing to market is very important. The easyJet model condenses this core timing issue in just a few variables and feedback loops allowing quick feasibility tests of the strategic initiative to complement managerial judgement and to challenge prevailing wisdom. This type of model rehearses the basic intuition of the manager in order to find out hidden pitfalls. It is also small enough to illustrate fundamental concepts in system dynamics such as stock accumulation and feedback loops and therefore serves a useful pedagogical purpose too.

In contrast Sterman’s People Express Management Flight Simulator, about the growth strategy of a US low-cost airline in the 1980s, is a much larger model of several hundred equations. It examines the problem of coordinating investment and hiring in a fast-growth no-frills airline (a service business) while maintaining staff motivation and high-quality service. The scope of the problem situation is defined more broadly than for easyJet and the dynamic phenomenon to be explained – growth and unintended collapse of the firm – requires a more sophisticated dynamic hypothesis and model.

2.3.4 Public Policy: A Medium-Sized Hospital Model

In 2004, the European Working Time Directive EUWTD, a measure intended to limit the working week to 48 h for all workers within the European Union (EU), became mandatory for junior doctors working for the National Health Service (NHS) in Britain. The intuition behind this piece of health and safety legislation was to reduce the fatigue experienced by junior doctors by limiting doctor’s working hours and so improve the quality of patient care. This rationale, though compelling, does not address the non-clinical effects of the EUWTD, in particular, the fundamental change that the directive has on doctors’ working patterns, in-service training and work-life balance.

These concerns were of personal and professional interest to Dr. Mark Ratnarajah, a paediatric specialist registrar based in London who, at the time, was also enrolled on the Executive MBA programme at London Business School. He decided to conduct a project to consider the effects of the directive on junior doctors’ career decisions and the consequences of these decisions on the medical workforce and quality of patient care. Based on his experience and knowledge of the UK National Health System NHS he developed a system dynamics model to consider the broad implications of the directive and to explore alternative courses of action.16

16 For more information about the model see Chapter 9 of Strategic Modelling and Business Dynamics.
The project was conducted in two stages. In stage one a workforce planning model was built to explore how hospitals will cope with the expected loss of junior doctor cover and the transition to full-shift work patterns. This model focuses on the tangible effects of the directive on the total hours available from junior doctors. In stage two the model was extended to include intangible effects of the working time directive on the work-life balance and morale of junior doctors and potential knock-on consequences to doctors quitting the medical profession. The models’ structure was derived from the modeller’s own decade of personal experience as a physician trained in the NHS. Parameters were gleaned from government healthcare policy documents and from journal articles about the medical profession.

2.3.5 Reflections on Model Fidelity and Size

These examples show that system dynamics models of varying fidelity can support the process of strategic development. There is no one perfect model of an organisation that will reveal the future outcome of strategy and policy with certainty. Modelling is fundamentally the art and science of interpreting complexity, and there is always a choice about how much detail to include, depending on the purpose. On the one hand there are small scale models, mere sketches of a complex reality, whose purpose is to reflect managerial intuition and rehearse the implications. On the other hand there are larger, more sophisticated models whose purpose is to facilitate strategic change by developing shared understanding of complex situations and by testing the effect of specific policies.

2.3.6 Required Skills of Practitioner

The skills needed to conduct projects depend on the problem situation, model size and whether or not a simulator is required. Someone who is familiar with the rules of causal loop diagramming (and who is confident with group facilitation) can conduct projects of similar scope to Bernhard Kerres’ study of Orchestra Management. However the expertise necessary to create clear and insightful loops should not be underestimated.

Simulators, both large and small, require skills in formulation, equation writing and simulation analysis. As the drug related crime model shows, there is a significant step in going from a causal loop diagram to a full-fledged simulator of the same feedback structure. Normally such work is done with project teams that include both policymakers and expert modellers/facilitators. The relevant modelling expertise can be found among members of the system dynamics community, in niche consulting companies, and among graduate students who have specialised in system dynamics. Useful gateways to members of this community are the websites of the System Dynamics Society www.systemdynamics.org and the Society’s UK Chapter www.systemdynamics.org.uk.

I should emphasise that it is not easy to build full-blown simulators and there is always room to improve modelling skills. Even small metaphorical simulators present
significant formulation challenges for the novice as the following story about the Romeo and Juliet model illustrates. The tiny model contains just two stocks: Romeo’s love for Juliet and Juliet’s love for Romeo. They are mutually dependent. Two strikingly simple connections are all that is needed to produce a cyclical pattern of love in the time chart. At least that is the dynamic hypothesis. If the change in Romeo’s love for Juliet depends on Juliet’s love for Romeo, and vice-versa then an endless cycle of waxing and waning love is possible. This hypothesis comes as a big surprise to many people.

From the stock and flow diagram it is a further step to a full-blown simulator. In my experience this step is not easy. In executive education programmes I sometimes ask participants to formulate equations themselves. First they write equations for the two stock and flow networks (copying the standard syntax that applies to all stock accumulations). Then they tackle the tricky task of formulating equations for causal links between the two lovers. I allow participants the freedom to introduce auxiliary concepts in order to operationalise the links. This exercise, conducted in pairs, provokes a lot of thinking and discussion. Participants try their best to capture the imagined sensitivity of lovers and argue whether Romeo and Juliet respond to being loved in exactly the same way or somehow mirror each other’s affections. The model is small enough that everyone manages to formulate a full-set of equations and run simulations. The result is a wide variety of time charts. Some charts show escalating growth of love while others show a collapse of love to a permanent state of cold lovelessness (zero units of love).

In the limited time available it is very rare indeed for anyone’s model to reproduce the intended cyclical pattern of love (although the exercise provokes much fruitful thought about the relationship between Romeo and Juliet, just as a metaphorical model should). Nevertheless participants learn a useful cautionary lesson. The exercise shows that it is difficult to write equations that mean what you intend (and of being absolutely clear about what you really mean). Herein lies an enduring challenge of good system dynamics modelling. The same challenge applies to business and public policy models, only more-so. It can take 2 or 3 days of the project team’s time to come-up with a conceptual model worthy of the problem situation. It can take weeks or months more of the modellers’ (or modelling team’s) time, depending on model size, to formulate equations and then create a calibrated and fully-tested simulator suitable for evaluating new policies and strategies. In other words, don’t expect a credible simulator to appear overnight. It is necessary to carefully work through all five iterative steps described earlier in order to build confidence in the model, its structure, equations and fitness for purpose.

2.3.7 Enhancing Your Skills in Feedback Systems Thinking and System Dynamics

This chapter on system dynamics is necessarily condensed. If you wish to learn more about the subject then there are several good sources to consult. My own book *Strategic Modelling and Business Dynamics* (Morecroft 2007) covers all stages of
model building from problem articulation to mapping, equation formulation and simulation. It includes a range of in-depth practical examples that vividly illustrate important or puzzling dynamics in business, society and everyday life. The book also includes software and simulators that allow readers to run models described in the text and to role-play in dynamically complex systems.

Another good source is *Business Dynamics* (Sterman 2000). This comprehensive and definitive textbook thoroughly explains the philosophy, theory and practice of system dynamics modelling and simulation. It is exceptionally well written and provides a wealth of case model examples from business and society. Although the book is used in advanced courses on system dynamics, several chapters are well-suited to beginners. For example, causal loop diagramming is covered in Chapter 5, with many well-documented examples. Then, in Chapters 6 and 7, there is an excellent treatment of stocks and flows and the dynamics of stock accumulation.

For those who are interested in strategy and system dynamics there is *Strategic Management Dynamics* (Warren 2008). This textbook provides the basis for an entire strategic management course based on sound dynamic principles of asset stock accumulation. It includes explanations of how these principles connect with many of the most widely used frameworks in the strategy field. There are also extensive worksheets and exercises to develop skills in mapping a firm’s strategic architecture in terms of interlocking tangible and intangible asset stocks. Simulations show how this architecture delivers performance through time.

A final suggestion is to sample a PhD dissertation with the intriguing title “How and Under What Conditions Clients Learn in System Dynamics Consulting Engagements” (Thompson 2009). The author tackles the important yet slippery topic of model-based learning drawing on his considerable experience as a businessman, keen observer of system dynamics, serious student in the field and system dynamics consultant. The work involves documented histories, from ten consulting engagements, of clients’ learning experiences in modelling projects they themselves initiated. A combination of direct observation, survey, interview and personal reflection provides compelling stories of client insights, significant learning events, ‘aha’ moments, and some setbacks. Among the client organisations are a pharmaceutical company, a medical care provider, a development bank, a medical insurer, a memory device firm, a community hospital, a shipyard and a manufacturer of steel balls or “boules” (used in the bowling sport of petanque). With applications spanning service and manufacturing, private and public sector, system dynamics is indeed a versatile systems approach to managing change.

References


Systems Approaches to Managing Change: A Practical Guide
Reynolds, M.; Holwell, S. (Eds.)
2010, XIV, 309 p. 112 illus., Softcover
ISBN: 978-1-84882-808-7