INCREMENTAL DECISIONS IN A COMPLEX WORLD

PROF CHRIS ELLIOTT[1]

Decision makers are used to working with problems that are ***complicated****, that is rich in* ***detail****. Whether they are trying to manage a computer or a farm, a football crowd or a multi-national company, managers know how to get the task done - break the problem down into its constituent parts, get an expert to solve each part and stitch the whole together with a management hierarchy.*

This approach falls down when applied to problems that are ***complex****, that is, rich in* ***structure****. What do you do when the decision taken in one area affects the behaviour of another? How do you decide to close an uneconomic railway service without understanding the impact on roads, housing or employment? How do you trade off speed, fuel economy and comfort when deciding which car to buy, or mass, radiation resistance, reliability and image quality when designing a weather satellite?*

These problems are the domain of system engineering. This discipline has grown out of defence and aerospace and can be applied to complex social, technical and economic challenges. System engineering seeks quantitative answers, not qualitative guessing; a description of the whole, not just the parts; answers that are good enough, not exact, and so offers the decision maker a rational basis for choice.

Introduction

In 1962 President Kennedy set the American nation the challenge of 'putting a man on the Moon and bringing him back safely before the end of the decade'. In 1969 Neil Armstrong said (or should have said) 'that's one small step for a man but a giant leap for mankind'. Between those two speeches tens of thousands of people made millions of individual decisions. In the video film with which this lecture opens one of them says 'I don't know how all of this mission works, I don't even know how most of it works but I do know how my bit of it works and it ain't gonna fail because of me'.

This lecture is about taking decisions. There are two problems - how do you take the individual decision that faces you when you do not have sufficient time and resources and how do you ensure that your decision fits together with all of the decisions that your colleagues are taking? The second problem is one of systems so I'd better start by defining a system. A system is a set of components which, when they are brought together, exhibit properties that were not present in the components alone. These properties are called emergent properties and they often make it difficult to predict how a system will behave by examining the components alone.

I shall be coming back to systems later in this lecture but let me illustrate an emergent property in action. Figure 1 shows a spring hanging from a support and joined by a short string to a second spring. A weight is hanging from the lower spring and loose strings are tied from the support to the top of the lower spring and from the bottom of the upper spring to the top of the weight.

---

[1] Chris Elliott is a Charter Professor of the Principles of Engineering Design at the University of Bristol, a consulting system engineer and a practising barrister.

Pitchill Consulting, Magalee, Moon Hall Road, Ewhurst, Surrey GU6 7NP (tel/fax +44 (0) 1483 273463)
Chambers of Marie-Claire Sparrow, 95A Chancery Lane, London WC2A 1DT (tel +44 (0) 171 405 3101)
chris.elliott@pitchill.demon.co.uk

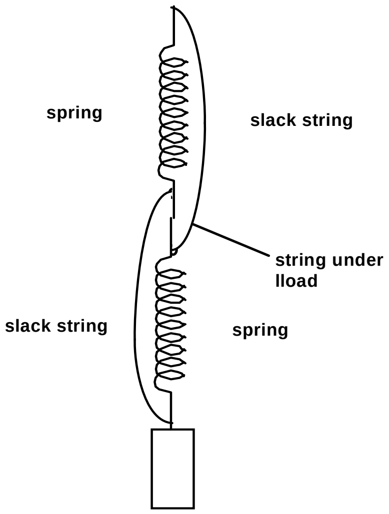


Fig 1

The question is: 'what will happen if I cut the short string joining the two springs?'. Clearly the weight will fall, but how far? Three possible options are shown below.

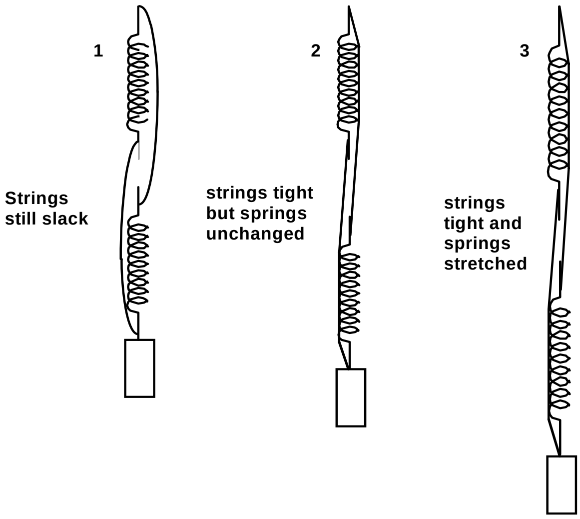


Fig 2

Position 2 is the point at which the two loose strings are just tight and the springs are unaltered. I think that it looks like the most likely result but not everyone agrees. Position 1 has no visible means of support and to me looks very implausible. Position 3 is lower than position 2 and requires the springs to stretch when the string is cut.

A show of hands around the audience at the lecture showed that the majority thought that the weight would fall to position 3 and a few people voted for position 2.

When the string is cut, the weight does not fall but **rises** by around 10 cm. On the face of it, this is remarkable since I had cut a string under tension and the two things that were pulling it apart became closer together. However, if we redraw the picture as shown in figure 3, we can see the reason why. Electrical engineers will see it immediately - the springs were 'in series' before and are 'in parallel' afterwards. When they are in series, each carries the full weight. When they are in parallel, each only has to carry half the weight and so shortens.

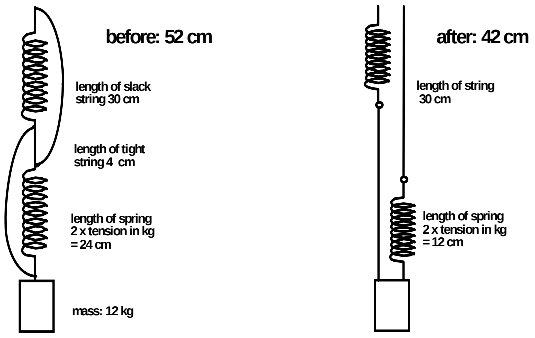


Fig 3

I have included some illustrative numbers for the dimensions and a simple Hooke's Law definition of the spring stiffness. There is no doubt - the weight rises. In fact, the behaviour is quite complicated and depends on the length of the strings, the mass of the weight and the stiffness of the springs. If we increase the mass, so that the springs stretch to take up the slack in the 'loose' strings, the weight still rises.

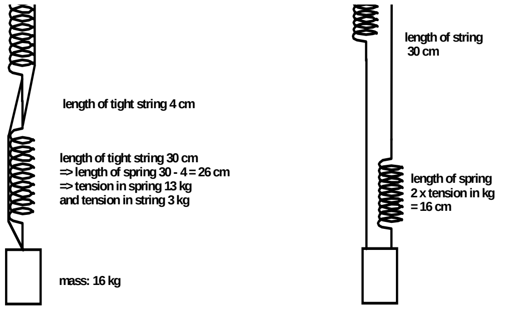


Fig 4

I shall come back to the properties of systems later in this lecture. Before that I want to look more generally at the world of engineering design. One of the greatest engineers of the twentieth century, Sir Ove Arup, said when delivering the Institution of Structural Engineers Maitland Lecture in 1968:

> “As in art, its problems are under-defined, there are many solutions, good, bad and indifferent. The art is, by a synthesis of ends and means, to arrive at a good solution. This is a creative activity, involving imagination, intuition and deliberate choice, for the possible solutions often vary in ways which cannot be compared by quantitative methods”

He was talking about engineering design and that is my main theme but the words that he used might equally be used for problems in politics, business or one's personal life. The engineer's dilemma can be illustrated easily - shall I use a 6 mm bolt or an 8 mm bolt? The former is lighter and cheaper but the latter is stronger and hence safer. There is no right answer, just as there is no right answer to the political decision on how much to spend on health, the business decision as to whether to invest in a new factory or the personal decision on where to take a holiday.

I shall start by considering what I call 'simple' decisions - those are the ones where you know what it is you are trying to achieve and the difficulty is achieving it (I've put simple in quotes because they are often far from simple in practice).

After that, I shall consider complicated decisions, where you are not even sure what a good solution looks like. Finally I shall consider complex decisions which involve systems and their emergent properties.

Simple decisions

Investment decisions

Let's start with a decision making problem that interests anyone who is paying into a pension fund - which investments should be made to maximise the value of the fund. Let us assume that ten times per year the manager invests five per cent of the fund in a speculative investment. We will also assume that half of speculative ventures rise in value by ten per cent more than inflation. The other half fall in value by the same amount.

Clearly if the fund manager selects his investments at random, without using his expertise in the markets, the average value of this fund exactly tracks the inflation index (if we ignore dealing costs). What would we expect to be the distribution of values of a large number of funds, each of which was managed according to this principle? After five years, the standard deviation will be ≅50 x 5% x 10% ~ 3.5%. We would therefore expect the value of the funds to follow the normal distribution curve as in figure 5, centred around the rate of inflation.

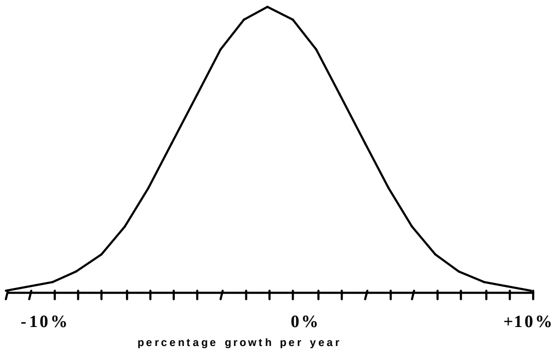


Fig 5

What happens in practice? I used the Internet to find the value of the 371 US growth funds that have been trading for five years. Figure 6 shows the histogram of those values, superimposed on the normal distribution of figure 5.

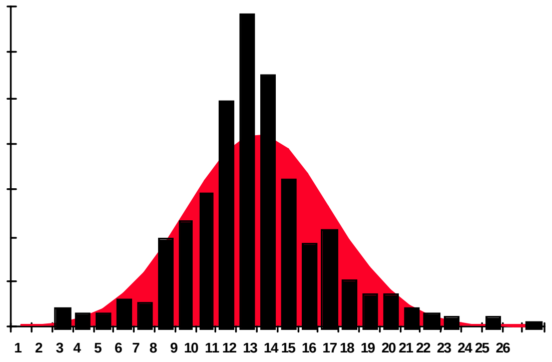


Fig 6

The mean growth rate is 12.3%, slightly less than the growth in the Dow Jones index over the same period which was 12.6%. The first observation is that the specialist fund managers have, on average, added no value by their efforts. They have just made enough to cover their dealing costs and fees. It is then interesting to note that more funds are grouped around the centre than the normal distribution would predict. This, together with the relatively small number of funds in the 'shoulders' of the distribution, indicates that many of the fund managers did not bother to trade as much as would be expected. More funds did much better than would be expected - perhaps they are the insider traders. Finally, more did much worse than would be predicted - perhaps they are incompetent inside traders.

What does this tell us? It is not just a chance for an engineer to make fun of City dealers - the point is that there are some decisions that do not merit a great deal of effort. Where we do not have enough information to take a rational decision, or where the consequences of our choice are not predictable, there is no point in putting a great deal of effort into the decision. This is a very important message for an engineering designer (or any other decision maker). Since you rarely have enough time to analyse all of the issues that are amenable to analysis, don't waste time on those that are not.

Detection - two ways of getting it wrong

A recent development by AEA Technology is a computer that recognises handwritten signatures. It works by calculating a score for the similarity between the signature and a reference version stored in its memory. If the score is above a threshold value, the computer says that the person making the signature is who he claims to be. If the score is below the threshold, it says that he is an impostor. The way in which we write our signatures varies so the system cannot be perfect - the score it will attribute varies around the correct value.

'Countermatch' has been successfully used by the Employment Service, the Ministry of Defence and by banks. Each of them had to take a difficult decision - at what level do I set the threshold? The Employment Service is concerned not to reject any valid applicant since this would be very embarrassing. It therefore sets the threshold quite low to avoid any false rejections. The Ministry of Defence, on the other hand, is concerned that noone should gain unauthorised access to sensitive information and sets the threshold very high. It is prepared to annoy a few genuine users to be safe.

The banks are even more subtle. If you want to withdraw £10, they set the threshold at a low value since it is better to make a few incorrect payments than to annoy customers. If on the other hand you try to withdraw £10 000, the threshold is set at a high level to ensure that they are paying the correct person.

It follows that there are two kinds of mistake:

- falsely accepting an incorrect signature
- falsely rejecting a valid signature

and the user can bias the decision-making system so as to make one less likely at the expense of increasing the probability of the other. This allows us to introduce the precautionary principle in decision making. You can estimate the consequences of each kind of mistake and bias the decision making process towards the kind of mistake that is less damaging.

This kind of decision making was first subject to rigorous mathematical analysis by the early developers of radar who called it "detection theory". Radar works by transmitting a powerful pulse of radio energy in a narrow beam. The pulse is scattered by any obstruction that it encounters and some of that scattered energy reaches the receiver of the radar set. The strength of that 'echo' is an indication of the size of the obstruction and the time between sending the pulse and receiving the echo is proportional to its range. A further complication is that the further away the obstruction, the weaker the echo.

Now let's imagine the radar on a warship. The pulse is transmitted over the sea and is scattered by the waves and any other obstruction, be it a seagull or an Exocet missile. The challenge is to decide when the received energy is an echo from an Exocet, and to decide in time to take evasive action.

When the Exocet is near, the signal is very strong. If we plot the received energy over time, it might look like figure 7. The normal signal, labelled 'noise', is a combination of the echoes off waves and other benign obstructions and the inherent electronic noise of the receiver.

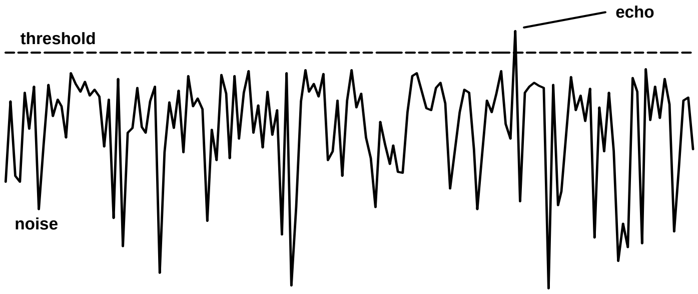


Fig 7

We can set a threshold that will reliably detect the echo and hence know that the missile is present. If however we want to detect it at greater range, the threshold will have to be lower and we will start to get signals crossing the threshold which are not missile echos. Once again, we have the choice of two sorts of error (called false alarms and missed detections).

Where do we set the threshold? If it is too high we will not detect a missile in time to react. If it is too low, we will use up all of the ammunition in our Close In Weapon System before a real missile is ever encountered.

False alarms are like the boy who cried wolf and devalue the subsequent real detections. Decision makers, being mindful of the precautionary principle, have to weigh up the probability and consequences of each kind of error when setting the threshold.

Trends and fluctuations

We can also apply the ideas of decision theory to slowly varying phenomena, where the challenge is sometimes that of 'seeing the wood for the trees'. Let's take global climate change as an example. The change is slow and the fluctuations from year to year can mask the underlying trend. One of the techniques used by climate scientists is to look at historical data over a long period.

There are many sets of data that one could use. An example is the average temperature in December in England. Figure 8 shows that temperature every ten years since 1683.

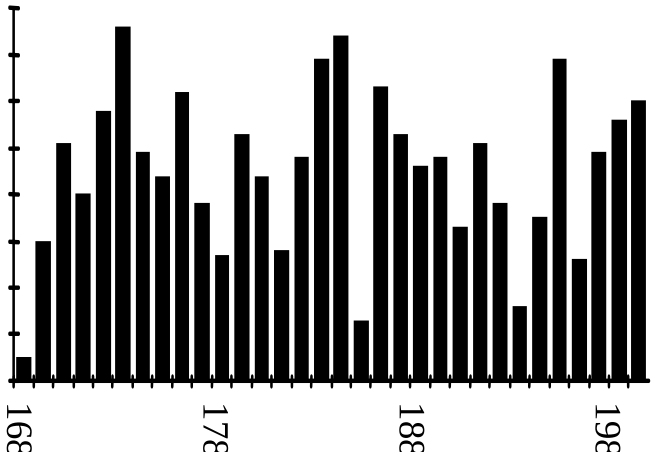


Fig 8

It is hard to detect any pattern or systematic change. If we start to calculate running averages, a trend begins to emerge. Fig 9 shows the running average over fifty years, superimposed on the raw data. The increase over the last ~ 100 years is beginning to emerge.

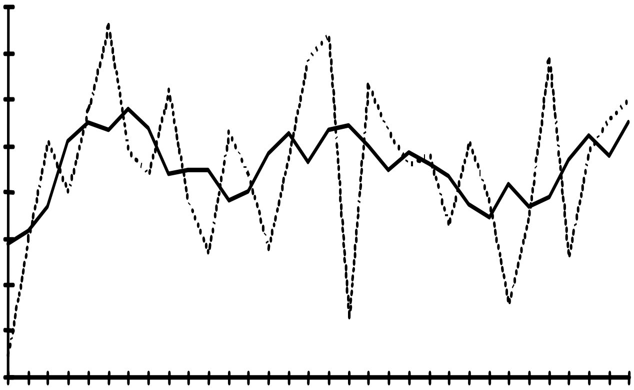


Fig 9

Figure 10 shows two more running averages of increasing length. The solid line shows a clear trend, with a warm period from around 1750 to 1850, a cooling in the late Victorian times and a warming during this century. The dotted line shows that too much averaging smooths not only the fluctuations but also the underlying phenomenon that we are trying to detect.

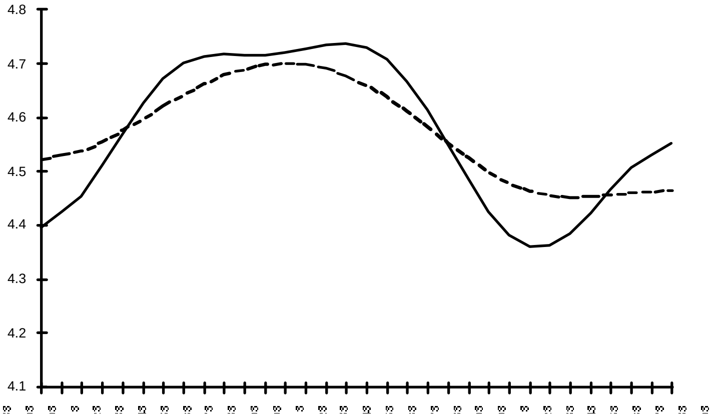


Fig 10

Now the question is whether the trend that we see in this century is a cyclic variation or the beginning of a more serious effect. Clearly there is much more data available but again we have to use the precautionary principle. If we decide that it does indicate a real problem and we are wrong, we will waste money and effort on changing the way we generate energy and run society. If we decide that there is no problem and we are wrong, we leave an even greater problem for our grandchildren.

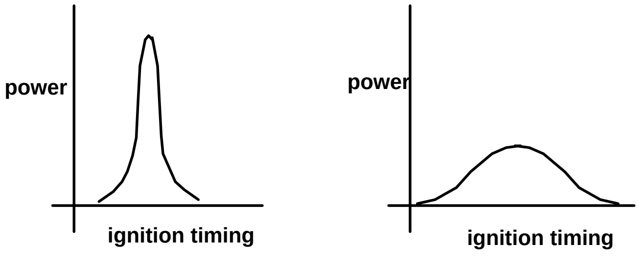
Conclusion

This section of the lecture concerned 'simple' decisions. I hope that it is clear that actually taking the decision is often far from simple but at least we know what it is we are trying to do - detect the impostor, detect the missile or detect climate change. Even that is not clear for complicated decisions.

Complicated decisions

Optimisation - what is a good solution?

Let's take a simple optimisation problem - the power produced by a car engine as the ignition timing is varied. There is a setting of the timing which gives the greatest power and any change, either advance or retard, reduces the power. When it is correctly set, a racing car engine produces much more power than a road car engine but the performance peak is much narrower.



racing car **road car**

Fig 11

Which is better? If you want to win a race, you have to have the power of a racing car engine at its peak. If you do not have that power, one of the other competitors who does will beat you. However, the setting is very sensitive and any slight change will drastically cut the power available. If on the other hand you want to drive to work, reliability is more important than peak power and the road car performance curve is much safer. The answer to the question 'which is better?' is 'it depends on what you want to do'.

Another transport example - most people would say that a thoroughbred race horse is better than a camel but is that true when you are halfway across the desert? - more about camels later. First it is necessary to look in more detail at 'measures of goodness'.

It is necessary to introduce two definitions:
efficiency - how well does the solution use resources?
effectiveness - how well does the solution meet the needs?

Consider a train pulling a load up a hill. The measure of effectiveness is the rate at which it raises cargo up the hill, which is the rate of climbing times the mass of cargo. What fraction of the total mass of the train should be allocated to the engine?

It is reasonable to assume that the power of a train engine is proportional to its mass. The rate at which it can lift mass up a hill is proportional to the power of the engine divided by the mass of the train, which equals the engine fraction. The rate of raising cargo is the rate of climbing times the mass of cargo, and thus is proportional to the engine fraction times (1 - engine fraction). Figure 12 shows this measure of effectiveness plotted as a function of engine fraction.

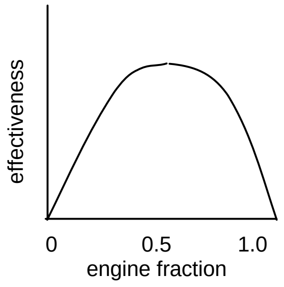


Fig 12

It is clear that the train is most **effective** when it is only fifty per cent **efficient**, that is, where half of the total mass of the train is engine.

A better known example of exactly the same effect is the electrical circuit shown in figure 12. If the aim is the heat the water as quickly as possible, the resistance of the heater, r, should be equal to the internal resistance of the power supply, R, even though half of the energy is then dissipated in the internal resistance.

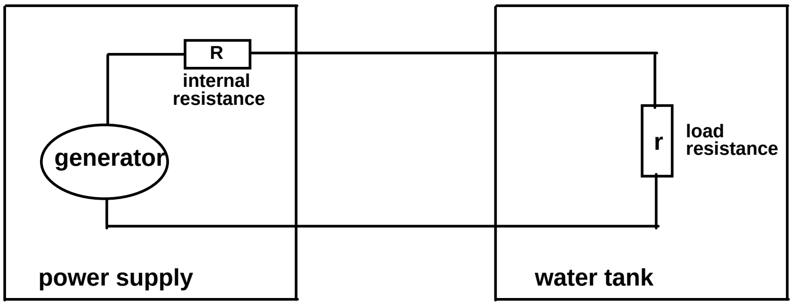


Fig 13

A more subtle example can be found in nature. Photosynthesis is the process by which plants convert light energy into stored chemical energy. It is a seven stage process and around 0.8 per cent of the incident energy ends up being stored. If each stage is fifty per cent efficient, the overall efficiency would be expected to be $(0.5)^7$, which is almost exactly 0.8 per cent. It looks as though evolution has, not surprisingly, resulted in optimum **effectiveness**, rather than high efficiency.

Measures of effectiveness

Unfortunately engineers are rarely allowed the luxury afforded to evolution. They are supposed to build one solution and get it right, not build many different solutions out of which the most effective survives. They need a measure which allows the potential effectiveness of each possible solution to be predicted before it is built. Any real problem has many qualities which have to be considered and we have to combine them to calculate a score.

Let's start with a non-engineering example - which university is best? 'The Times Good University Guide' takes ten separate qualities and estimates how well each university scores on each measure. It then calculates a combined score for each university by adding together each of the individual scores. This is a way of bringing together what could be called 'hard' qualities like the number of Firsts awarded or the staff/student ratio and 'soft' qualities like the standard of student accommodation.

The result that emerges is not to everyone's liking. The Chairman of the Committee of Vice Chancellors and Principles said 'The compiler has taken single characteristics and combined them in a highly specific way that will be irrelevant to the needs of many students'. An individual student might have particular needs or interest - sporting facilities, wheelchair access, specialisation in a subject - and the combination of scores will not reflect that student's requirements.

The general measure of goodness may suit many students but not all. This is a reflection of the more general observation that there is no such thing as absolute quality - only fitness for purpose. The measure of goodness that is specific to the needs of a particular customer or user is usually called a Figure of Merit.

In its simplest form a weight is attached to each quality to reflect its importance to the particular customer or user. The Figure of Merit of a candidate solution is found by multiplying the score achieved for that quality by the weight attributed by the user and summing over all qualities. 'The

Times' used a weight of one for each of the qualities in its calculation. More complicated Figures of Merit can be defined, including:

- hard thresholds - 'I won't consider any university that does not score above 75 on student accommodation'

- combinations of qualities - 'I want a university with a high number of postgraduate students and excellent research, so I'll multiply those two scores together rather than add them'

- non-linear rules to prevent a single quality preempting the result - 'Any points over 90 for staff/student ratio will add half as much to the overall score as points below 90'.

The idea of a Figure of Merit is simple but its implications are far reaching. The crucial point is that it makes the decision making process into a one dimensional problem. All the time that you have two separate measure of goodness, you have no rational basis for making a choice. Purists are uncomfortable with the idea of combining dissimilar qualities but there is no choice. Any choice has to weigh up the relative merits of the different qualities - a Figure of Merit is a way of making that process objective and visible.

Once a Figure of Merit has been defined, it becomes a design tool. The engineer can estimate the score of each of the candidate solution before building them. More importantly, it becomes possible to make a rational trade-off between conflicting qualities. One example of a trade-off is between price, performance and schedule. It is usually possible to satisfy any two of these but only at the expense of the third. So, an engineer could build the solution that does everything you want and could deliver it when you want it, but the cost will be excessive. If you insist on reducing the price, either the quality or delivery date will have to slip.

A more contentious trade-off involves safety, cost and performance. Statements like 'safety is paramount' are meaningless. A completely safe train would never leave the station. Instead we have to use a Figure of Merit to trade between these three qualities. For example, the UK Department of Transport uses a rule that it is worth spending around £750 000 on road improvements to save one life.

Trade-off and decision making are at the heart of engineering design. There is no 'right' answer - the aim of the designer is to find the least bad compromise.

Systematic design

The principles of decision making that I have developed now allow us to challenge Sir Ove Arup's claim that possible solutions cannot be compared by quantitative methods. We can apply those principles by setting out to solve any design problem by answering three questions:

- what is the problem?
- what solutions are possible?
- which is the least bad?

What is the problem?

This might seem to be the easiest question to answer but often is the hardest. A useful guiding principle is 'the customer is rarely right'. Customers frequently pose their requirements in terms of prospective solutions and it is part of the designers' job to find the true underlying need. A famous example will illustrate this.

During the Apollo moon landing programme, it was realised that the astronauts would need to write notes and that a normal ball point pen needs gravity to make the ink flow. NASA invested millions of dollars in developing a pressurised pen that would work in free-fall. Many years later, in an early period of *detente*, an Apollo capsule linked to a Russian Soyuz capsule in orbit. The cosmonauts came drifting through the hatch carrying not only a sociable bottle of vodka but also a pencil.

The mistake was to specify that what was required was a ball point pen that would work without gravity, rather than a writing implement that would work without gravity. The lesson is that the engineer can only define the correct Figure of Merit after a thorough analysis of the customer's real needs. In particular, he has to identify what qualities the customer requires and what weights to give them.

What solutions exist?

In many ways this is the easiest question, not because it is intrinsically easy but because it is what everyone recognises as design. Research and development feeds into the inventive and innovative process of devising candidate solutions. I only want to draw attention to one danger which brings us back to the subject of camels.

We considered earlier the problem of optimisation, which I illustrated as a hump - the performance of the solution varies smoothly as some parameter is changed and the aim of the designer is to get to the top of the hump. That is fine if one is engaged in dromedary engineering, where there is only one hump. However, many engineering problems are bactrian - they have two humps.

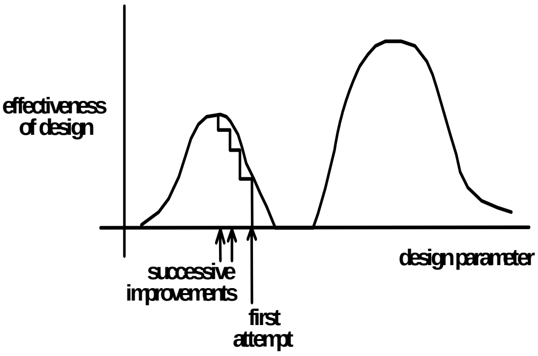


Fig 14

Engineers are generally very skilled at making incremental improvements to an existing design. They are less good at finding the correct starting point. Figure 14 might be thought to be describing the evolution of personal computers - the first attempt was DOS and the successive improvements were Windows and Windows 95. Meanwhile, the Mac alternative was there all the time.

A current engineering problem is that of the electric car. Huge efforts are going into developing ever better batteries. The essential quality of a battery is that it converts chemical energy into electrical energy and is recharged by reversing the process - electricity is pushed back into the battery which is converted into stored chemical energy. The alternative is a fuel cell which also converts chemical energy to electrical energy but is recharged by adding new chemicals. The potential performance of the fuel cells corresponds to the hump on the right of figure 14 but many researchers are struggling on with small improvements to batteries, corresponding to the hump on the left.

It should now be even clearer how the ball point pen could happen. If the problem is incorrectly defined, it will exclude the solution on the right without it even being considered.

Which solution is least bad?

At this point the designer's job gets easy. Question 1 gave him a Figure of Merit, Question 2 gave him a range of candidate solutions and all that is left is to calculate the effectiveness of each candidate and select the least bad.

Two points are worth noting:

- if one solution is obviously much worse than another, do not waste time trying to work out exactly how much worse - save the effort for the viable candidates that need more careful analysis.
- if two candidate solutions have similar scores, do not let the Figure of Merit determine the choice. It is generally a 'rough and ready' tool and not accurate enough for fine decisions. The precautionary principle is more important - 'how could this solution fail' - so always think about the fall back position if something goes wrong.

Systematic design in practice

I'll illustrate this with a real but greatly simplified system problem to show how these three questions are tackled in practice. The task is "build an environmentally-friendly transport system". The example shows how far you can get with just a good model of the problem, some simple physics and a few parameters (obtained by two phone calls, to friends at MIRA (Motor Industry Research Association) and the former British Rail Research).

What is the problem?

There are many specific questions to be answered to define the problem:

- what is to be transported? (eg people, bulk goods, parcels)
- how frequent are the journeys? are they regular or random?
- how long are the journeys?
- what are the environmental effects that must be considered?
- what infrastructure already exists?
- how is the system to be financed?

In order to keep this example simple, I shall assume that the general answer to Question 1 is "a commuter transport system that uses the minimum energy". I'll make the following operational assumptions:

- the number of passengers is predictable (a commuter service will have similar demand on successive days);

- the maximum waiting time must not exceed 15 minutes (and hence that a vehicle is needed every 15 minutes, whether it is full or not);

- every passenger must have a seat;

- the vehicles will travel at 80 km per hour.

It is necessary to define a Figure of Merit to measure the effectiveness of each solution. I shall use energy consumption per passenger kilometre and shall assume that:

- energy is needed to accelerate the vehicle and to overcome aerodynamic drag and rolling resistance;

- the vehicles travel at a constant speed and acceleration and deceleration times are negligible.

The total energy used in a journey of the *d* km with *n* passengers, *E* (kJoule), is given by:

$$E \quad = \quad (M + nm)\, V^2/2 \quad + \quad (M + nm)\, Rd \quad + \quad r\, c_d\, d\, AV^2/2$$

where
M = mass of vehicle (kg)
m = average mass of a passenger (assumed to be 80kg with luggage)
V = speed (ms $^{-1}$)
R = rolling resistance (N tonne $^{-1}$)
r = density of air (kgm $^{-3}$)
c_d = drag coefficient
A = frontal area (m^{2})

and the energy per passenger km is E/nd.

What solutions are possible?

Again to be simple, let's assume that the solutions of interest from Question 2 are single deck bus, double deck bus, conventional train and light railway. It is now necessary to choose between them.

The relevant properties of the four types of vehicle are:

		bus single deck	bus double deck	train 3 unit	train light rail
mass of vehicle	tonne	8	12	111	40
rolling resistance	N/tonne	60	60	6.4	6.4
drag coefficient		0.55	0.55	0.8	0.5
frontal area	m2	7	10.5	9	9
no of passengers (seated)		35	70	216	60
no of passengers (total)		35	100	280	250

Which solution is the least bad?

I'll only consider a single leg journey for convenience. The equations and data that I stated above are sufficient to work out the absolute energy consumption per passenger km as a function of the journey length by each kind of vehicle, provided that one knows the number of passengers, the vehicle speed and the journey length.

Now we have to start considering the operational assumptions. Figure 15 shows the best (least bad) solution as a function of the number of passengers per hour and the journey length:

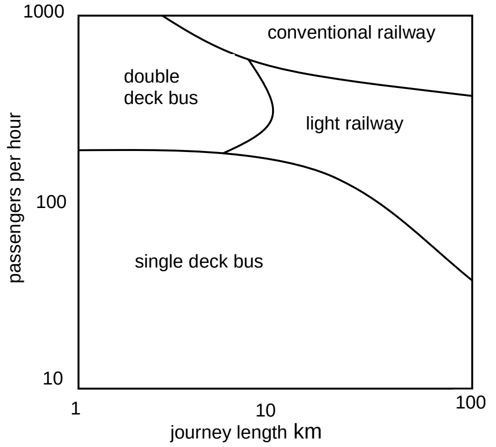


Fig 15

This result shows clearly how the “best” solution is critically dependent on the exact requirements of the service. How robust are the answers? What happens if we assume, for example, that speed will be halved, to 40 kph (25mph)?

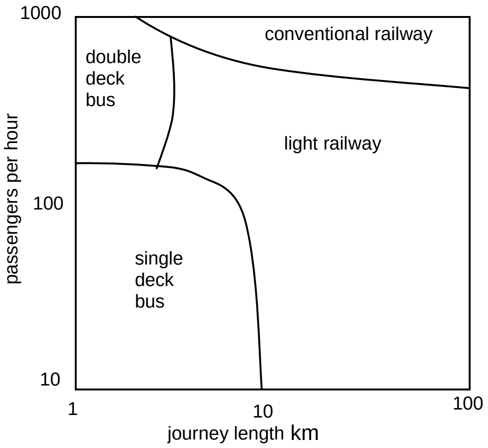


Fig 16

Light rail has a greatly increased range of passenger density and journey length for which it is the preferable but the general form of the picture is similar to that at the higher speed. Figure 17 shows what happens if we make another change to the assumptions, by including standing passengers.

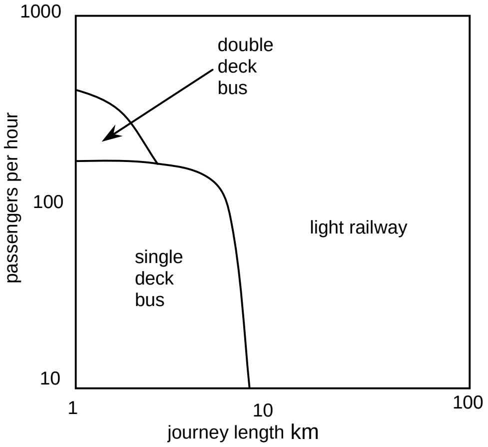


Fig 17

It would appear that light rail has “stolen” all of the territory of conventional rail and much of that of buses. Even so, it appears that buses are more energy efficient for urban journeys, as well as arguably being cheaper and more flexible. I leave it to you to decide whether standing accommodation for a 100 km journey is an acceptable service.

More generally, a few simple calculations based on readily available data have provided a rational basis for selecting the most energy-efficient means of transport. Did anyone do those calculations before building the Manchester Metro or the Croydon Tramway?

Detailed design

The process that I have been describing is the one that is needed before starting detailed design. Its sole aim is to find the general form of the least bad solution and leaves the different and in many ways much harder job, requiring different talents and skills, of implementing the design. However, it is a job for which there are many well-developed design tools, from Computer Aided Design and simulation through to modelling and field trials.

I hold those who can do that job in the greatest respect since I know that I could not do it well but, unless the system design is right, they will end up with a beautifully engineered wrong solution. The American writer Robert Benchley summed up this argument very well when he said that he thought that the most difficult part of building a bridge must be starting it. 'I might be able to finish it if someone would start it for me, but as for the first move, I would be left blushing furiously'.

Complex decisions

Systems in practice

The principles that help when taking simple decisions (the two sorts of mistake, the precautionary principle) and complicated decisions (what is the problem?, what solutions exist?, which is the least bad?) can be applied in the context of complex decisions about systems - a set of components which, when they are brought together, exhibit properties that were not present in the components alone. Now we have to consider that the decision that you take about one of the elements may have an impact on the emergent properties and thus depends on bigger issues than the element alone.

Let's start with a simple system. The components are a battery, a light bulb, a red wire with clips on its ends and a similar black wire. If I clip one end of the red wire onto one terminal of the battery and the other end onto one of the tags of the light bulb, and clip the black wire onto the other battery terminal and the other tag of the light bulb, the bulb glows. The emergent property is light, which was not there in the components before they were brought together.

Now if I repeat that demonstration using a different black wire, nothing happens. This 'wire' is not a wire at all, it is a piece of black plastic tubing. It is necessary to understand how the system works **as a system** in order to select the right components. The second 'wire' looked like the real thing but lacked a vital but invisible ingredient - a copper core.

During the Second World War, the natives of Papua New Guinea were fascinated by the behaviour of visiting troops. These strangers levelled a section of forest and painted some white lines on the ground then machines landed from the sky and unloaded beer and cigarettes. After the war ended, the 'Cargo Cult' persisted in preparing airfields but, because they had no understanding of system of which airfields were part, were unable to understand why no beer and cigarettes arrived.

It is not necessary to go so far from home to find the 'Cargo Cult' mentality. A British Home Secretary announced that part of a policy on law and order would be the introduction of mandatory life sentences on a second conviction for rape. Wise judges immediately pointed out that there can be no sentence without a conviction and, if the only possible sentence is life, juries will be reluctant to return guilty verdicts unless the evidence is absolutely certain. The result will be that on average rapists will receive lighter, not more severe sentences. A further twist is that, since the sentence will be the same for rape and murder, a rapist might as well kill his victim to prevent her identifying him.

The judicial system has many elements of which statutory sentencing policy is only one. The Home Secretary had, like me using the black 'wire' or the Cargo Cultists building runways, latched on to one element of the problem without considering the system as a whole.

Similar examples abound. A very simple one occurred recently in environmental legislation. Toxic zinc waste used to be recycled and used in new products. It was then declared to be a dangerous substance so, instead of being recycled, must be dumped in landfill sites thus increasing the environmental damage - the exact opposite of the intention of the legislation.

Systems are like snakes - if you grab hold of the tail, the head twists round and bites you.

One more general example will serve as a useful introduction to the more detailed case studies that follow. This comes from farming. The infectious disease sheep scab has been largely eliminated by the compulsory dipping of sheep. All sheep were dipped and there was no opportunity for the disease to persist. Dipping was then made optional. Each farmer had to trade off the costs of dipping against the risk of infection (a simple decision by my definition). Even following the precautionary principle, most would correctly come to the conclusion that it was no longer worth dipping. The result was the reappearance of sheep scab because there was now a large number of undipped sheep in which it could thrive.

The significance of this example is that each farmer's decision not to dip was correct, given the costs and risks involved. It is only when all of the farms are considered together that the problem arises.

Transport systems

The map in figure 18 shows a commuter choice. There are two roads from Suburbia to The City, each passing through a village. The road from A to B and the road from C to D are fast dual carriageways and the average journey time is 30 minutes. The roads from A to C and from B to D pass through the villages and are congested. This means that the time taken to drive them depends on the number of cars using them. Assume that the time taken is 2 minutes for every car per minute so, if 5 cars per minute are going through a village, the time taken is 10 minutes.

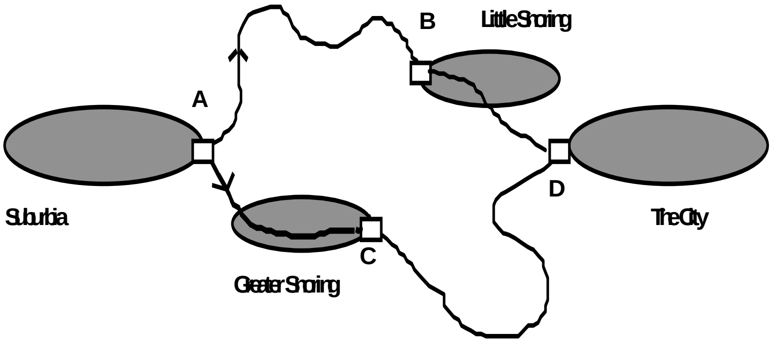


Fig 18

Now we will assume that 16 cars per minute travel from Suburbia to The City at rush hour. After a few days, the drivers will learn which route is quickest and the traffic flow will settle down to 8 cars per minute using each route. The journey time for each driver will be (2 x 8) + 30 = 46 minutes. This is a state of equilibrium, where no driver can reduce his journey time by changing his action (what economists call a Pareto optimum).

Now let's assume that we build the Snorings By-pass, from B to C to reduce the traffic through the villages (fig 19). The by-pass is also a dual carriageway and uncongested, taking 4 minutes to drive. On the first day that it is open, let's assume that one driver per minute who reaches C takes the bypass. He has driven through Greater Snoring (16 minutes), along the by-pass (4 minutes) and through Little Snoring (18 minutes since there is now 9 cars per minute through that village). His journey time is 38 minutes, significantly shorter than before. The others at C who go straight on to D still take 46 minutes but those who went from A to B now take 48 minutes because of the extra car's contribution to congestion in Little Snoring.

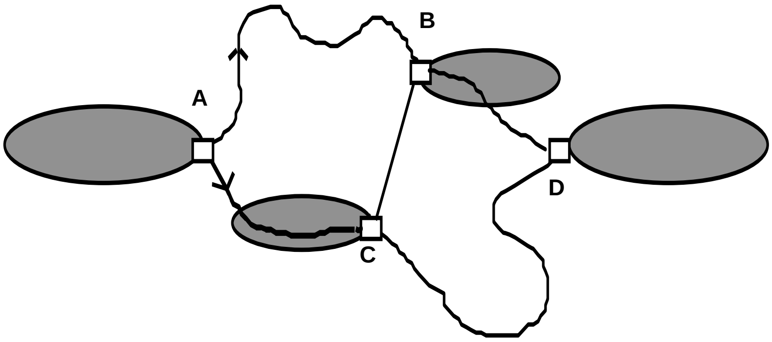


Fig 19

It is still worth taking the bypass from C to B and more cars will do it. Eventually, when equilibrium is reached, 13 cars per minute go through each village and each journey takes 56 minutes. The result of the by-pass is a 60% **increase** in the traffic through each village and a 10 minute **increase** in everyone's journey time.

Is it realistic? It has been suggested that this might be a consequence of the Newbury by-pass. This phenomenon was first predicted in the 1960s by Braess, a German mathematician, and some of his colleagues claim to have observed it in Stuttgart as early as 1968. You might also recognise the numbers. They appeared at the start of this letter in order to explain the surprising behaviour of the springs when I cut the string. The equations governing the road network are identical to those governing the springs. Cutting the string is equivalent to closing the by-pass.

Incidentally, you will probably realise that I chose the numbers in these examples carefully. Once the rush hour traffic rate exceeds 30 cars per hour, the by-pass does become effective. Similarly, if I had hung a greater weight on the springs, it would have fallen when I cut the string. The messages that I am trying present are:

- that you need to analyse a system carefully before making a change to be sure that your change will have the effect that you desire; and

- the free market in which every player takes the decision which is best for him locally (be he a driver or a sheep farmer) may give a result in which everyone loses.

Now lets look at a public transport system. Figure 20 shows a rail network consisting of six stations along a line. I shall assume that 1000 people per day wish to travel between any two stations (so 1000 travel A to B, another 1000 travel A to C, yet another 1000 travel B to C and so on) and that the fare that they will pay corresponds to £1 per stop for the return journey. I shall also assume that the cost of running each section of track is £3000 per day, except the section from C to D which costs £10 000 because of its difficult construction. The total income earned by the network is £35 000 per day and the total running cost is £22 000 per day - clearly a viable proposition.

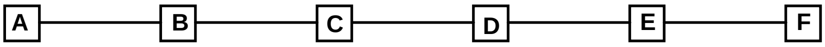


Fig 20

Now what happens if we split up the network and make each section self-accounting? Section A to B and E to F have an income of £5000 per day and cost £3000 - they are still viable. Sections B to C and D to E have an income of £8000 per day and a cost of £3000 - they are even better. Section C to D has the highest income, £9000 per day, but costs £10 000 so is uneconomic. If it has to be truly self-accounting, it must close.

Now what happens to the other sections? Their passenger numbers all fall and the income is only £2000 per day each. Since each costs £3000 per day to run, they too must close. The previously viable network has collapsed because of incremental accounting decisions. This has nothing to do with ownership. If the sections were each privately owned and operated, the owners of the outer sections could get together and subsidise the operator of C to D for their mutual benefit but that requires one to analyse the network as a system. Piecemeal analysis is no different from my black 'wire' or the Cargo Cultists' airfields.

In conclusion

I have tried to argue that there are simple decisions which have two sorts of error and you can choose which you prefer, complicated decisions where it is not even obvious what a good solution looks like, and complex decisions where the perverse behaviour of systems means that your decisions can only be made in the light of many other decisions.

This is the world inhabited by the design engineer. I started with a quotation from an eminent structural engineer and tried to show that perhaps he was unduly pessimistic. I shall finish with a quotation from another eminent structural engineer, but this is one that I cannot fault:

> ***“Engineering is the art of moulding materials we do not wholly understand into shapes we cannot precisely analyse, so as to withstand forces that we cannot really assess, in such a way that the community at large has no reason to suspect the extent of our ignorance”***
>
> Dr A R Dykes, Institution of Structural Engineers, 1976