Chapter 3
The Viable System Model[1]

Patrick Hoverstadt

Abstract The Viable System Model (VSM) is a conceptual model which is built from the axioms, principles, and laws of viable organisation. It is concerned with the dynamic structure that determines the adaptive connectivity of the parts of the organisation or organism; what it is that enables it to adapt and survive in a changing environment. It can be used as a comparison against an actual organisation in order to identify weaknesses, mismatches or missing elements in diagnosing a problem and then as a framework for organisation design to resolve a diagnosed problem. Also it can be used for purposes of design from a clean-sheet. At the foundation of the model is the concept of variety, the number of possible activities of the parts and the necessity to limit these to those required for survival. The breakthrough in developing the model was the understanding that this could only be achieved with a fractal (recursive) layered structure. Furthermore at each level the pattern of the regulation of the variety of possible activities must be fractal. The chapter takes the reader through the development of the model and shows how the VSM supports autonomy and adaptablility.

This chapter, written by a highly experienced practitioner, Patrick Hoverstadt, describes the model and its elements from a practitioner perspective supported by practical advice and helpful recommendations on its use.

[1]This chapter uses edited excerpts and selected figures from Hoverstadt (2008) *The Fractal Organization: Creating Sustainable Organizations with the Viable System Model*, copyright of John Wiley & Sons Limited.

P. Hoverstadt (✉)
Fractal Consulting
e-mail: patrick@fractal-consulting.com

M. Reynolds and S. Holwell (eds.), *Systems Approaches to Managing Change: A Practical Guide*, DOI 10.1007/978-1-84882-809-4_3, © The Open University 2010.
Published in Association with Springer-Verlag London Limited

3.1 Introduction

3.1.1 What Is VSM and What's It for?

In the 1950s Stafford Beer was a senior manager in a steel company and began to develop new thinking in management by drawing on his understanding of control systems as described by the then new science of cybernetics and on systems theory, particularly from the fields of social research and biology. The complete VSM model was first published in 1972 in 'Brain of the Firm' where he first set out the development of the model through an application of cybernetic principles to the functioning of the human body. When he developed the Viable System Model (VSM), Stafford Beer was seeking to develop a "science of organisation", using systems and cybernetic principles that underpin all organisations (Beer 1959, 1966, 1974, 1978, 1979, 1981, 1985, 1994). His criterion was how organisations create viability, which is the capacity to exist and thrive in sometimes unpredictable and turbulent environments. This requires that organisations are or become ultra-stable, that is capable of adapting appropriately to their chosen environment, or adapting their environment to suit themselves, even if they find themselves in a situation that has not been foreseen. This doesn't just mean that we are looking at a system to fulfil some given or ascribed purpose, we are also looking at how systems create their own purposes and maintain or change those through time.

I'm going to concentrate on the use of VSM to model human activity systems, though as I mention in the "reflections" at the end, there are a number of other uses it can be put to and is put to. In using VSM with Human Activity System organisations as the term is commonly used, there are three principle uses: diagnosis, design and self knowledge. Diagnosis and design are fairly self explanatory. In diagnosis, the modeller uses VSM as a normative model to compare against the real world situation to look for weaknesses, mismatches or missing systemic elements that explain the problem being experienced or at least give a handle on it. Design can either be a clean sheet exercise (let's sit down and design this new organisation), or following on from diagnosis (let's redesign this part of the organisation to deal with the problem). The third common use comes from Conant–Ashby Theorem (Conant and Ashby 1970), one of the basic tenets of systems and part of the internal logic of the VSM. Conant–Ashby says that "every good regulator of a system must be a model of that system" – in other words, your ability to manage an organisation depends on how good your model of that organisation is. Overwhelmingly, the most common organisational model in use is the hierarchical model. Hierarchy is originally a religious concept and is about "nearness to god". The fundamental belief is that the higher up you are, the closer you are to infallibility. In practice, what a hierarchical model actually models is the overt power structure or more prosaically, the blame structure. It doesn't model a number of quite important things you need to know to understand an organisation, such as: what it does, how it does it, how and where performance is managed, how the parts are coordinated, how the organisation adapts, how or where it takes decisions, and on what information those decisions are taken. All of which the VSM does cover. The reason I came to use VSM was simply that it allowed me to understand how organisations work when they do and why they don't work when they don't – far better than

anything else I had come across. And so far, I haven't found anything else that comes close to it in dealing with problems to do with organisation.

3.1.2 Overview of the Model

The VSM is presented as a graphical model – a picture with a number of critical components (five sub-systems and an environment) that are connected together in a particular way and are needed for viability (see Fig. 3.1). The subsystems are:

- System 1 – the set of activities that the organisation does which provide value to its external environment, the primary operations (System 1 is drawn in the standard diagram below as a set of circles)
- System 2 – the set of activities or protocols to coordinate operations that are needed to stop the different operations causing problems for one another (represented by the triangles on the right hand side of the diagram)
- System 3 – the management activities to do with allocating resources to operations and ensuring they deliver the performance the organisation needs, which we might call 'managing delivery'
- System 4 – the management activities to do with understanding the environment and the future, with planning and change, the outcome of which is to develop the organisation
- System 5 – the set of management activities to do with ensuring that the organisation works as a system, specifically that there is a balance in decision making between Systems 3 and 4, and also maintains the organisation's identity and ensures that activities undertaken are consistent with acceptable practice, what we would normally call governance.

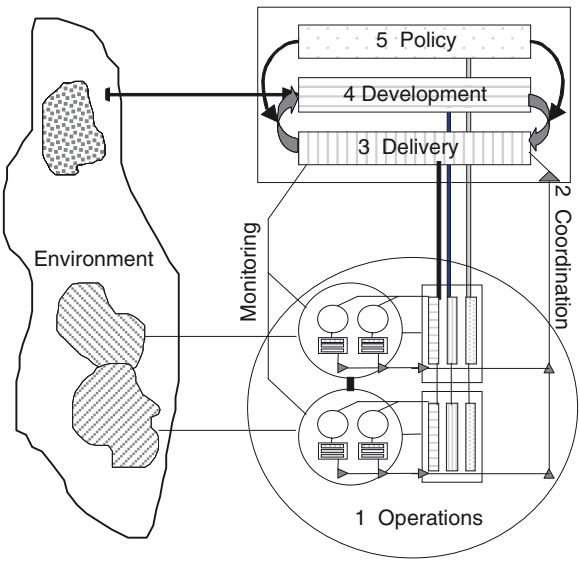

Fig. 3.1 Viable systems model

- The environment is modelled as outside the system in focus and conventionally is represented as an amorphous blob.

I've gone into these sub-systems in more depth in the following sections. The five subsystems are different types of activity that are connected together in a particular way. They aren't necessarily different people or teams or departments. Particularly in a small organisation one person can be performing several of the functions identified above and can be active in several areas of the model if they do different types of activity. Conversely, one circle on the diagram might represent the activity of a whole division of a multi-national organisation. So the model is fundamentally different to a conventional organisation chart in that it represents types of activity rather than "things".

3.1.3 Key Concepts

Beyond this basic graphical model which shows the static systemic structure of the organisation, broken down by type of activity and by the connections between those activities, there are a number of key concepts that we need to address. Mostly these have to do with complexity and the ways in which VSM handles complexity.

The basic VSM model with its five subsystems is fairly simple, how then to deal with a large complex organisation? VSM does this by being a recursive or fractal model. This means that within the "operations" circle of System 1, there will be a set of operational sub-activities, each of which will also be a viable system with exactly the same systemic needs and systemic structure as the whole. So, we have viable systems made up of viable systems which are made up of viable systems and all of which use the same systemic architecture (see Fig. 3.2).

So, a team needs to manage its resources and performance and change, just as the department it's part of does, just as the division the department sits in does, just as the corporation the division is part of does. In practical terms then, we can use a relatively simple model to deal with organisations of any degree of complexity.

The connections in the VSM diagram are just shown as lines, but they actually represent two way communication channels and "variety equations". The VSM is a working through of Ashby's Law of Requisite Variety (Ashby 1956) which says that "only variety can absorb variety", where variety is a measure of complexity – "the number of possible states of the system". What this means is that if we have an environment that demands six varieties of service or product from us and we can deliver all six, then we have "requisite variety", whereas if we can only deliver five we don't have "requisite variety". Which seems pretty obvious – about as obvious as Newton's observation that apples fall from trees – and in management terms about as significant as Newton's Law of gravity was to physics. Why? Because, the environment our system sits in and relates to has much higher variety than the organisation's operations and the operations have higher variety than management, so the question of how to balance these inherently unbalanced variety equations so that the organisation can be managed to carry on delivering what the environment needs is a non-trivial one and the fundamental problem that VSM sets out to address.

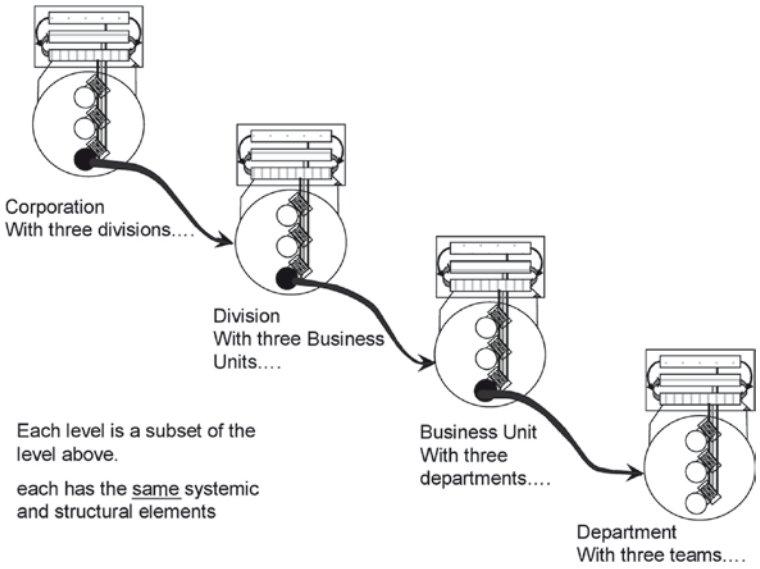

Fig. 3.2 A fractal structure

This problem of balancing variety equations which are intrinsically weighted against the management of the organisation drives the two most critical tensions in the VSM: the tensions between the autonomy of the parts versus the cohesion of the whole and the tension between current delivery and future need. Both of these have their own section in what follows.

They are also intensely relevant to two of the fundamental concepts in systems generally: wholeness and emergence. The original basic tenet of systems approaches is that there are attributes – emergent properties – that systems have as a whole that they do not have as components. So, they can only really be understood as cohesive wholes. The autonomy – cohesion tension within VSM is about this. Too much autonomy and all cohesion – the wholeness is lost. Too much cohesion, too little autonomy and emergence is reduced and "wholeness" is impoverished. VSM provides a language to debate how to set this critical balance.

3.2 The Model: Underpinning Concepts, Structure, and Use

3.2.1 Autonomy Versus Control

3.2.1.1 The Horns of the Dilemma

There are few issues in management that are quite as contentious, quite as likely to trigger strong emotional reactions as the question of authority and autonomy. Even within the same organisation, you can find managers who argue passionately that

centralised control by a hierarchy is critical and next to them managers who are equally passionate that centralised control dooms organisations to fatal rigidity in a fast changing world. The two sides often caricature one another. The advocates of hierarchy complain about anarchists and the advocates of greater autonomy depict the supporters of hierarchy as slightly sinister control freaks.

The passion betrays the underlying fears on both sides of the debate. Both sides know that the other's arguments have some validity, but aren't completely right. Organisations that are too centralised are too rigid, do find it difficult to adapt to changes in their environment and do die as a result. Organisations that have no centralising decision making structures are incapable of acting as coherent wholes and do fall apart. The problems are real. The dilemma is real and part of the reason for the emotions is that many managers recognise that they are caught in a dilemma – which is not a comfortable position to be in.

3.2.1.2 The Complexity Equation

When Henry Ford started production of the Model T Ford, the world's first mass produced car, he was famously reported as saying that his customers could have it "any colour – so long as it's black". His manufacturing philosophy was in line with Frederick Taylor (Taylor 1911) the great advocate of management control. Taylor reasoned that one of the principal roles of management was to control work practices to reduce proliferating variety. Following the Taylorist line, several generations of managers sought to set down and control how staff did their job, sometimes in great detail.

For many years now, Taylorism has been seen as outdated, as an approach that inhibits change and innovation. For me, the interesting question, and one that is frequently ignored is "what has changed?" For make no mistake, if Taylorism is rightly seen nowadays as generally being an unhelpful approach in today's environment, it wasn't always so. Time was when it worked and worked well. The stunning success of Henry Ford's Model T – 15 million were made between 1908 and the late 1920s at a time when most other makers' models were produced in hundreds or fewer – proves just how successful the Taylor approach was. So if it did work once and doesn't now, why is that? What has changed? The answer is two things, one external to organisations and one internal.

Internally what changed to make the Taylorist formula redundant was increased complexity of technology and skills. At the beginning of the twentieth century, it has been estimated that 95% of workers couldn't do their job as well as their immediate boss. At the beginning of the twenty-first century, it is estimated that this statistic has pretty much reversed, so that 95% of workers can do their job better than their boss. A century ago, when a factory needed to appoint a new supervisor for a machine shop, they would simply promote the best machine operator working in the shop and they would become the new supervisor. Because the most skilled were promoted, of course they could do the job better than their staff. In that context, the Taylorist approach of managers dictating not just what was to be done, but how it was to be done made sense.

Nowadays, it is normally the case that staff understand how to do their job better than their bosses and management is seen as a separate skill-set in its own right, not just something that the best operators will acquire through osmosis. In this context, the idea that managers can centrally control all aspects of operations is simply nonsensical and the level of autonomy of staff has to be radically different from the Taylorist model.

Externally what has changed is the complexity of the environment we operate in. No car company these days could realistically survive, never mind prosper to become the biggest car manufacturer in the world if it was only prepared to offer cars in one single colour. A market that Henry Ford was able to treat as if it was largely homogenous has become progressively more and more segmented and fragmented – more complex. Henry Ford's dream was to bring car production to a position where it could create a new mass market. Whilst other producers were hand crafting individual commissions at luxury prices, the Model T was designed and built by semi-skilled workers and was sold at a price to compete with horse drawn buggies. The market accepted the Model T as a basic no frills product because customers were new to the car market and had low expectations. So out in the environment, the market was simple for the Model T and Ford was keen to keep it that way, hence "any colour – so long as it's black". The problem that Ford did face in his environment was sheer volume, how to build something as complicated as a car in millions, not tens or hundreds. The answer was in the simplification and standardisation of the production process – the Taylor solution.

What Ford created was a balanced equation: outside, a simple undifferentiated market demanding high volume and inside a standardised process capable of producing standard products in high volume. The internal organisation was able to match the complexity of needs of the market by treating customers as essentially the same and offering a simplified product in great numbers. Where there were differences in customer needs, these were not addressed by Ford. They were dealt with by a huge sub-industry that sprung up to service, maintain and customise the basic car. For Ford, business success came from getting the right balance of complexity either side of the equation between the company and its environment.

Of course this equation wasn't stable through time. Increasing customer diversity between customer groups increased the complexity of the market. With the Model T, Ford had effectively been able to ignore differences between customers (other than geographic ones), but as the market matured, customers increasingly wanted not just a basic machine, but one that was suited to their needs and their tastes. To address this emerging problem, Alfred Sloan (Sloan 1962) developed the divisional organisation model used by General Motors. This brought in an organisational structural for GM that had specific units within GM each with its own branding and tasked with servicing a specific market segment. The increased complexity of the market environment was matched by a corresponding increase in the complexity of the organisation and so the equation between operations and environment was balanced once again. To do it, Sloan had to develop new managerial practices. These were designed specifically to cope with the autonomy divisions needed to cope with their different markets. The divisional management structure

allowed a degree of autonomy for divisions whilst still retaining overall cohesion. So as well as the complexity equation between environment and operations being in balance, the complexity equation between management and operations was also re-balanced. The formula was successful and propelled GM to become the biggest car manufacturer in the world (Fig. 3.3).

The next revolution came with the creation of the Toyota Production System (Monden 1983; Liker 2003) and here again there was an increase in autonomy to deal with an increase in complexity and now Toyota has taken over from GM to become the biggest car manufacturer in the world.

In the development of the car industry from 1908 we can see three huge shifts in organisational model. In each case, the change was designed to balance the fundamental problem of matching environmental complexity with an adequate operational response that could cope with the complexity of market demands. At the same time, increasing operational complexity demanded an increase in management response and this response was in the form of increased autonomy. The problem the industry faced was a simple problem of balancing complexity using Ashby's Law of Requisite Variety, which simply states that "only variety can absorb variety" which means that complex environments need organisations that are sufficiently complex to match those environments, and organisational complexity needs to be matched by management. Failing to match environmental complexity means that organisations fail to meet what the world demands of them and fail. Failing to match organisational complexity means that management cannot manage effectively, takes arbitrary decisions and fails. The problem is that simple. The same fundamental dynamic that has driven the development of the car industry affects every organisation of every size and in every sector. Every organisation faces the challenge of matching environmental complexity.

The trouble is that the complexity of the environment is theoretically infinite, so we have to be selective as to which aspects of the environment we are bothered about. Similarly, the organisation is more complex than management.

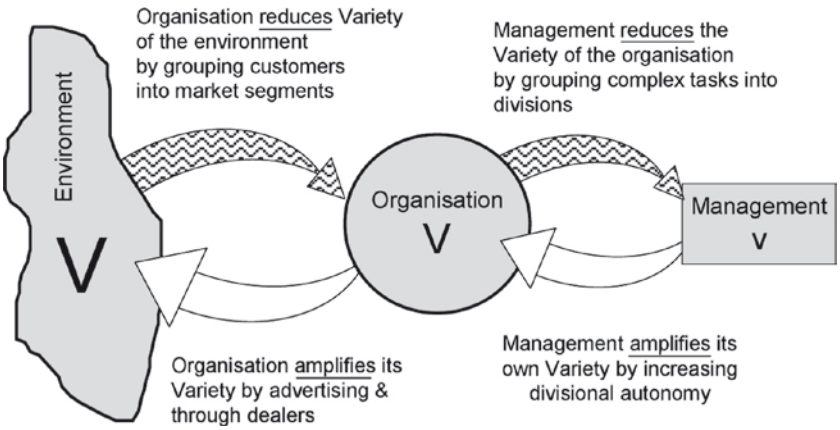

Fig. 3.3 Balancing the complexity equation

Reconciling what is a fundamental set of imbalances is what the VSM is all about. The balance can only be achieved by amplifying management's variety and attenuating that of the organisation and by amplifying the response of the organisation to the environment whilst attenuating environmental variety. Typical attenuators are to standardise and group. So we group individual customers into market segments and the organisation treats them as if they were the same. Similarly management groups complex tasks into divisions and departments and treats them as production systems with common reporting standards, not as individual tasks. Typical amplifiers include advertising to the market, but the most important is probably increasing the autonomy of operational units to address differences in demand. Understanding the level of environmental complexity that needs to be absorbed gives us a practical metric – admittedly a fairly crude one – for understanding the degree of autonomy we need for any organisation. The tension between sub-system autonomy and system cohesion is one of the most important tensions in the VSM.

3.2.1.3 Recognising Autonomy

One of the problems with hierarchy is that it is often an illusion. Even if you take an extremely coercive system such as a prison, where it would appear that the prison staff have huge power over prisoners, the reality is that the system can only function on a consensual basis. If the prisoners really decide they aren't going to play the game, then the system breaks down very quickly. This is even more true in more ordinary organisations where the apparent power of the hierarchy is very often more illusion – or at least consensual, than real.

In a large service organisation, the senior executive team operated a tight control regime. All decision making was centralised including detailed operational and resourcing decisions. There was absolute control of processes in the best Taylorist fashion, with detailed descriptions of how every aspect of operations was supposed to be carried out. Some of the executive team and senior and middle management argued they needed to get away from this "command and control culture". But, you didn't have to look very far before you came across lots of examples of staff ignoring the rules to ensure that the job got done. Overwhelmingly, when staff could see that the prescribed process was dysfunctional and where they could, they exercised the autonomy which was officially denied to them and went outside the official process. Procedures were regularly ignored and processes changed, steps omitted and others introduced. All this was done despite management decree. This wasn't a culture of "command and control", it was a culture of "command and ignore". The senior managers responsible for laying down the processes were blissfully unaware that middle managers were routinely taking control of their own processes. It was all done with the best of intentions and for the benefit of customers.

Leave aside any moral questions about the rights or wrongs of managers wresting control of their processes from senior managers, these managers were exercising their autonomy. They weren't gifted this autonomy. They weren't "empowered" to do it. It wasn't sanctioned. They just did it because they thought it was the right thing to do.

They had the power to do it and their bosses didn't actually have the power to stop them – because they didn't know it was happening.

This was Ashby's law at work again – as inexorable as the law of gravity. There was a mismatch between the complexity of the operational situations these managers were confronting and the responses provided by the officially endorsed processes. So given spare management decision-making capacity in the form of a bunch of smart well educated middle managers, they filled the vacuum and exercised their autonomy. People have autonomy to act whether we like it or not. As managers, we can choose to utilise that capacity, or to try to stifle it, but it exists and when we try to restrict it too much, it will find other outlets.

3.2.1.4 The Resolution of the Dilemma

So what's the difference between a hierarchy and VSM as far as autonomy and control are concerned? The fundamental difference is that in the VSM, it is clear that different levels of the organisation deal with different aspects – different types of complexity. This means that as the organisation is built up from its basic operations, there is a clear focus for management decision making at each level, and generally it isn't about the same things as at the level above or below. This gives a clear marker as to what management at each level should be focused on and what they are equipped to take decisions about and just as importantly, what they aren't competent to take decisions about. This is quite different from a hierarchy where the assumption is that senior managers know more than juniors about everything – down to knowing more about how to shovel coal into a boiler than the guy doing the shovelling. Using VSM, managers at different levels see different issues in the complex world they manage (both organisation and environment) from those at other levels. This means there is a need to have conversations between levels about how to proceed, if decisions at one level are not to destabilise decisions at another level. This doesn't imply that one level is subservient to another, since each is, or needs to be the expert in their particular environment.

The hierarchical model is about power. About who has the power to take decisions and it carries with it the assumption that higher in the hierarchy means better equipped to decide. The VSM is about managing complexity and difference and it carries the assumption that different managers in different parts of the organisation will be best placed to take decisions about their part of the organisation. Neither hierarchy nor anarchy, VSM provides a solution to the perennial debate about autonomy and hierarchy.

Many people have come to the study of VSM with one of two preconceptions. The first is that it is a hierarchical model and it has been severely criticised for this – quite unjustifiably. The second preconception is the exact opposite; that this is a model for organisation without control – almost an anarchist's charter. Both views are wrong. Viability demands that organisations have the capacity to balance the demands of their environment – which in complex environments rules out centralised hierarchies but it also demands systems that can act coherently so as to be effective – which rules out anarchy.

So our attack on the variety problem, requires that management at a particular recursive level agree with its operational subsystems a set of frameworks within which the operational subsystems should operate. The frameworks can only be obtained and maintained by agreement since the knowledge and expertise necessary to manage the whole system lies both in the management and in the operational subsystems. To set up and maintain this set of frameworks is the purpose of system two.

3.2.2 The Structure of Value Creation: System 1

3.2.2.1 Primary and Support Activities

Organisations are difficult things to build and run so there has to be a good reason for having one. The main reason for having an organisation is to do things that an individual cannot do on their own because the task is too complex. Either it is too big, requires more diverse skills than that individual has, or it needs to be carried out in several different places or at different times. In other words organisations are simply a way of coping with different types of complexity.

One of the critical steps in modelling an organisation either for design or analysis is to understand the structure of how the organisation deals with the complexity of the tasks it carries out. When I say tasks here, I'm referring specifically to "primary activities". These are the tasks that the organisation does that deliver value to the external "customers" of the system and I'm specifically not referring to all the tasks the organisation has to do to keep itself in being. In VSM, this is a vital distinction and however we choose to define identity, the distinction between primary and support activities is at the heart of understanding identity – of understanding "what business are we in" (Hoverstadt 2008; Beer 1985). There are different ways of distinguishing between primary and support functions, but the definition I use is based on the concept that there is some sort of value exchange between an organisation and its environment that keeps the organisation in being and that the activities that deliver this value are primary.

As an example, if we take the task of doing accounts in a building contractor, this is not a primary activity. It isn't the accounts that deliver value to the builder's customers. What they value is the building work the company does. By contrast, if we take the task of doing building maintenance in a firm of accountants, the building work isn't a primary activity, whilst doing accounts for customers is primary, because that is the service that external customers value. This distinction between primary and support activities is roughly analogous to the distinction of profit and cost centres in management accounting.

The term primary is a statement of the purpose the organisation exists to fulfil and the expectations that customers have of the organisation. It isn't a comment on the importance of tasks. Doing the accounts in the building company may be vitally important to ensuring that the company stays in existence and is able to service its clients, just as maintaining the building may be equally vital to the firm of accountants.

3.2.2.2 Organisation Structure and Complexity Drivers

So, starting with the primary activities of the organisation, the next question is "what is the best way of structuring these?" Each primary activity is made up of other sub-activities which in turn are made up of sub-sub-activities and we can decompose the tasks as far as we need to go to understand it. Building houses may be a primary activity of our building contractor, and that might be split down by building site, by individual building plot, by the different trades involved. If we wanted, we can carry on the task decomposition to the point where we are focused on the task of laying an individual brick, or knocking in a nail. Similarly with the firm of accountants, we could split the task up by specialism: tax, audit, management accounting etc. We can split the task up by sector, by customer, by geographic area: the London office or the New York office and just like the building company, we can carry on breaking down the task to the point where we focus on an individual calculation or check carried out. Since the organisation exists to do tasks more complex, more diverse in terms of skills, geography or time than an individual can cope with, the way primary activities break down level by level reflects the sort of complexity the organisation is trying to address. There are four principal drivers of complexity in primary activities (Espejo & Harnden 1989):

- Technology
- Geography
- Customers
- Time

"Technology" is about doing different things, so plumbing is a different job to bricklaying in the building firm and auditing is a different job to personal tax advice in accountancy – these are "technology" differences.

"Geography" is about structuring the organisation according geographic differences: different teams working on different building sites, or in different offices in the accountants.

"Customers" fairly obviously is about structuring activities according customers, so our accountancy firm might have a team specially set up to deal with big accounts and keep that quite separate from the team dealing with small clients. The builders might have a team dealing exclusively with "executive developments". In both cases, the rationale might be the specialist skills required for those sorts of customers.

"Time" is about continuing the job beyond the staying power of the individual or single team. So the most common example is shift systems in manufacturing or in 24 h services such as the emergency services, but it can take many forms such as having a duty officer to deal with "out of hours" emergencies.

Whatever the drivers of complexity, in analyzing any primary activity, it is important to realise that we are simply repeating the analysis process in unpacking the complexity. The resulting layered structure of system, sub-systems, and sub-sub-systems, etc. is called a recursive structure. The word 'recursive' indicates that the structure has the same pattern and properties at each level.

3.2.2.3 The Impact of Complexity Drivers

Primary activities are broken down into sub-activities according to one of these four drivers at each level. The order in which this is done – in other words the order in which the organisation's structure unfolds the complexity that it faces, can have an absolutely massive impact on how the organisation performs.

Let's take as a hypothetical example a government's provision of roads. This might involve two activities – road construction, and roads maintenance, giving us two organisational units using the same technology and in the same geographical area and for the same customers. Most likely one road repair team and one construction team will not cover the whole country, it may only operate in a particular location, let us say Erehwon. So to cover the whole country, there may be many such units that are divided by geography, perhaps on a county basis, all contained within the "Roads" agency, and each in turn containing a road construction and a road maintenance unit.

The "Roads" agency will itself of course be a part of a larger public sector body, say "Transport Infrastructure". In this case, it will be just one of several units that may be differentiated on the basis of technology, so roads may be one agency, railways another, urban light railways another. In this scenario, the diagram shows how the provision of roads is structured from the level of central government to an individual road project, and most importantly, the way that the complexity of this provision has been handled (Fig. 3.4).

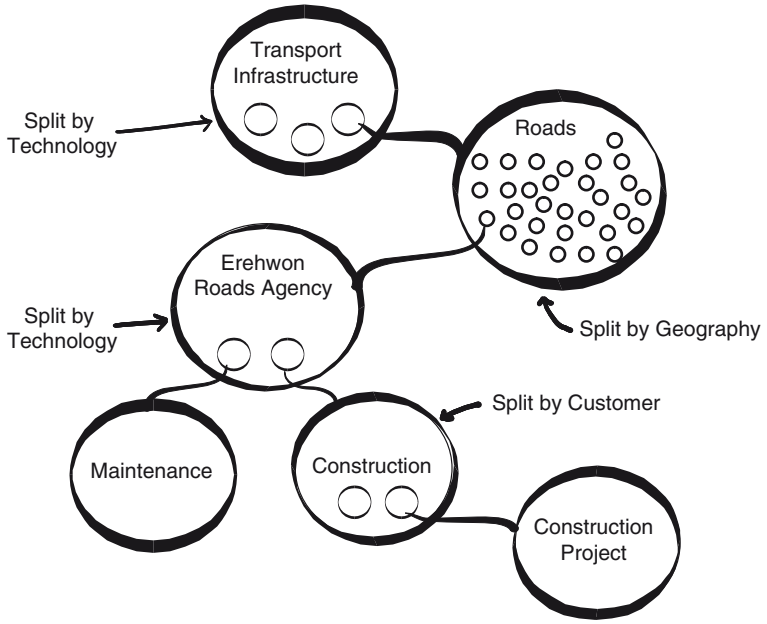

Fig. 3.4 Road Transport Organisation – Option 1

Although we have postulated this as a possible way of carrying out the structural division of transport infrastructure, it is by no means the only way of doing this. It could be done on a regional basis, with each county managing its own infrastructure: rail, roads, light rail, airports, etc. Or alternatively, it could be that regional division is done at the lowest level, and that all road infrastructure, both construction and repair is centrally controlled. A model for this might look like Fig. 3.5.

The critical issue is that the provision of roads to all areas of the country is a complex task, and the way that this complexity is dealt with has profound implications for the way that the organisation operates and the way that it is managed.

For example, in the first model in which we postulated an Erehwon Roads Agency that handled both maintenance and construction, we can easily imagine that it would be possible for the two to coordinate resource usage and swap both personnel and plant as needed. The implications of this may be a more efficient use of resources, but a drop in the speed of response of the road repairs service when maintenance resources were committed to construction.

In contrast, such a pooling of resources would be near impossible using the second model, since construction is controlled centrally, and only maintenance is managed at a local level. There are of course many other implications not only for the operations but also for the management. It is necessary to unfold the organisation's complexity in this way if we are to understand what these implications are for any organisation. In particular, this method allows us to start to look at where within an organisation decisions can be taken, and how resources may be allocated.

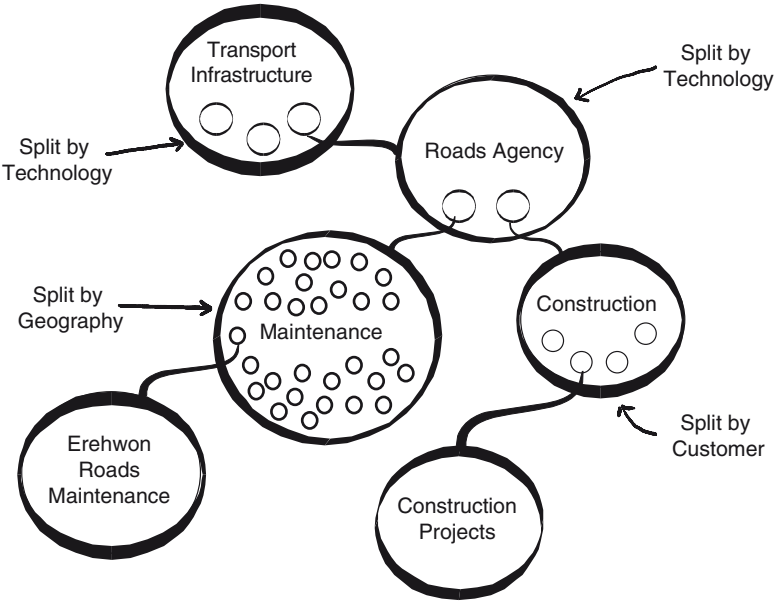

Fig. 3.5 Road Transport Organisation – Option 2

One common example of changing the order in which complexity is unfolded in an organisation's structure and the dramatic effects it can have is the switch in manufacturing organisations from functional departments to "cellular manufacturing". Back in the first half of the twentieth century, it was the norm that engineering factories were laid out in functional departments. So typically, there might be a turning department, full of lathes, a milling shop with milling machines, a drilling shop etc. Jobs would be passed back and forth between these shops having a series of separate operations done on them. This derived in part from the Tayloristic tradition of job specialisation and description which took task decomposition to extreme and assumed that restricting workers scope of work to a relatively few simple tasks would result in greater standardisation and improved productivity. Having worked on a production process with a cycle time of around 90 seconds I can vouch from personal experience that you do indeed get very good at doing it and you do get very fast, but it does get just a trifle boring.

The logic that this sort of task specialisation would be the most efficient seemed irrefutable, until firms started experimenting in the 1950s and 1960s with what was then variously called "group engineering" or as became more commonly know "cellular manufacturing". In this approach, groups of machines were put together in "cells". So rather than different types of machines being used in separate functional departments, there might be a mixture of lathes, millers, saws, drills etc. grouped together, with all the machines being used by a small team of multi-skilled operators. Each cell had the equipment necessary to carry out all the operations to make either a complete product or a complete sub-assembly that would go into a finished product. The results were dramatic. The accompanying table from a study of typical improvements with cellular manufacturing comes from a study done by London Business School of ten engineering companies (Table 3.1).

Don't forget, these improvements came simply from altering the structure of the organisation and therefore the way work was done. Some of the improvements are not too surprising. Reductions in work-in-progress, stocks and throughput time are easily accounted for; in functional departments, delays between operations in different departments are inevitable. A component would sit in one department while all the components in that batch were finished and would then wait (in some cases for days or weeks) till it was sent off to another department to have the next operation in the process done on it. By contrast, in a cell, as soon as each operation was carried out on a component it could be passed directly on to the next machine. In some cases,

Table 3.1 Improved performance from manufacturing cells

	Maximum %	Average %
Reduction of WIP	85	62
Reduction in stocks	44	42
Reduction in throughput time	97	70
Reduction overdue orders	85	82
Increase in sales	32	–
Increase in output per employee	50	33

process cycle times were slashed from weeks to minutes. This meant that at any time, there was drastically less WIP hanging round, fewer jobs in process (so less stock) but all of them moving very much faster.

Reduction in overdue orders is also easily explained, as production planning is very much easier and predictable. If a process that used to take several days because of all the delays built in now takes minutes, it becomes much easier both to accurately predict finish dates and also easier to push through a rush job.

Exactly the same design principles but applied to business process instead of manufacturing, formed the basis of the BPR revolution in the 1990s. Often the results were similarly dramatic, but often, the significance of the fact that what was now being streamlined were now often not primary, but support processes was lost. Very often, there were two ensuing problems: business processes that were hopelessly out of balance with the rest of the organisation and collateral damage to other processes as a result of not recognising the systemic role the process played. An Arthur D. Little survey of BPR initiatives found that of the successful ones, 68% threw up unforeseen harmful side effects.

3.2.2.4 Unpacking Complexity: Diagnosis and Design

Changing the order in which complexity drivers are addressed can change the organisation and its performance dramatically. Many corporate restructurings are about changing this order. Very often though, this is done without any clear rationale as to the relative benefits, or any method for working out why or indeed how one formulation will be better than another. The VSM provides a clear way of addressing this issue and a framework for working out the relative pros and cons of each structural option. There is never one single answer but in considering changes, we can be guided by the natural flow of the work. Each of the tasks we identify as a part of a primary activity is itself a primary activity. It will have its customers within the organisation.

Whenever we make decisions about how an organisation unpacks its complexity, this should be done by mapping this against the complexity of the environment and the complexity drivers operating there. But this isn't a static decision; each organisational response redraws the boundary between the organisation and its environment. When we do that, we can create or shut down opportunities. Each has its opportunities and dangers, but understanding what those are is critical to the decision. Changing the organisation to match unmet need in the environment – addressing a new or different complexity driver – has the effect of enlarging the organisation and changing the organisation's boundary with its environment. Changing the boundary means changing the organisation's exposure to its environment and so can lead to new opportunities or dangers. In health provision, research on new treatments which are intended to address unmet need often end up creating the possibility for yet more research into even more illnesses. Health provision is locked into a cycle of each new treatment creating the possibility for other new treatments, so the "market" for healthcare grows. This

is not necessarily a bad thing, but it is certainly a factor that needs to be considered when deciding on the organisation's basic operational structure. Some choices will expose the organisation to areas of the environment with many opportunities and dangers, others will offer far fewer.

In analysing an existing organisation, when looking at how the basic structure deals with the complexity drivers in the environment, as well as looking for the stress each option would put on System 2, we need to check how well each option addresses the complexity of the environment. Are we ignoring important distinctions between customers? Beyond the complexity drivers the organisation needs to address in the here and now, there is also the issue of what direction this will take the organisation in for the future, will it open up or close down future options. We like to think that we direct our organisations, and in a sense we do, but it is also true that our organisations circumscribe the sorts of strategy we are able to envisage and pursue. Our current decisions about how we deliver what we need today will largely determine how we relate to the world and that in turn will determine the future we are able to create. Mostly these choices are unconscious; they need to be conscious if we are not to have organisations that are simply driven by their history.

3.2.3 Maintaining Balance Between Primary Activities: System 2

3.2.3.1 Identifying Needs

We like to think of our organisations working as well oiled machines, where all the parts fit together, working in harmony with one another. Of course, it doesn't always work quite like that. Whenever we have a set of primary operational activities operating with any degree of autonomy, there is the possibility that one operation will do something that will disrupt the activities of another. The function of System 2 is to reduce or prevent inter-operation disruption (Hoverstadt 2008; Beer 1985).

The need for coordination increases with three factors:

1. The number of operational activities
2. The degree to which these can affect one another, or are interdependent
3. The degree to which they affect the same parts of the environment

The more integrated and more numerous our operations are, the more likely this sort of disruption becomes. The integration may be within the organisation, so if operation "A" supplies operation "B" they need to be coordinated. Equally, the connection can be through the environment. If two departments of the same organisation compete for the same customer, or send contradictory messages to the same market, that's a coordination issue. To prevent this sort of internally generated disruption we need some form of coordination between the operational activities at each level of recursion.

An extreme example of coordination problems was a large teaching hospital. With 60 service delivery units, there were too many different disciplines for

practitioners to understand what all the other departments were doing. This might not have mattered, if the care each offered was a discrete care pathway, but of course, because they were treating patients, they were related. Patients were no respecters of clinical boundaries. The patient who had come in with a broken hip was the same patient as suffered with Parkinson's disease and dementia, was malnourished and was in the process of getting bedsores. In this sort of situation, coordination problems go way beyond purely administrative issues such as having common standards for patient's notes. Different care needs can conflict and so need some way of sorting the prioritisation of clinical needs. Similarly, for a patient presenting at a hospital with a complaint that cannot be easily diagnosed, coordinating different disciplines to get the right specialist to correctly diagnose and prescribe the appropriate care pathway can be a very hit and miss affair.

Coordination problems have many symptoms that help in identifying them once they're happening:

- Oscillations in performance – the "shock wave" problem
- Low level ongoing chaos
- Cyclical recurring problems in operations – having to solve the same problem repeatedly
- Turf wars and inter-team or interdepartmental disputes

These are all classic indicators of missing or failing coordination. Of course it is always better to identify potential problems before they happen, so look for where there are connections between operational units, either where these are interdependent, or need to be but aren't.

Coordination problems rarely go away on their own. They tend to either occur periodically, simmer away constantly under the radar of management or are escalated to higher management for resolution. When this happens, they often trigger the "control dilemma" which can in turn jeopardise management including threatening strategy, so what appear to be low level and even insignificant operational issues can have a damaging effect on the organisation at a strategic level (Fig. 3.6).

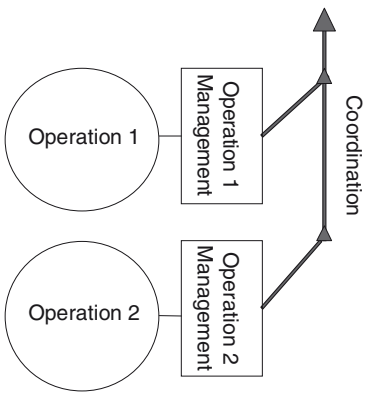

Fig. 3.6 System 2 coordination mechanisms to reduce disturbances between operations

Whenever organisations change, there will be a shift in coordination needs and addressing these will be a critical success factor for achieving change. Where they aren't addressed, they can prevent change happening and in a high proportion of change projects failure to plan new coordination is a key cause of failure. If department "A" is trying to change, but is also dependent on department "B", and there is no way of handling new aspects of their interdependence, managers are faced with changing and risking a breakdown of delivery or of staying with the status quo. Almost invariably given this choice managers opt for the status quo and change programmes stop. So anticipating coordination needs is important for both smoothing operations in the present and for enabling future change.

3.2.3.2 Coordination Mechanisms

Coordination failure, or rather the absence of coordination mechanisms, is one of the three most common systemic problems we see in analysing organisations. Generally, coordination is taken for granted when it is effective and is not correctly identified as the problem when things go wrong. It isn't as glamorous as heroic fire fighting for managers, but it is vastly more effective. We tend to praise and reward problem solving in organisations, but far more powerful than problem solving is problem anticipation and avoidance and this is what coordination does. The reason we take it for granted is because good coordination is so much a part of the infrastructure that we hardly notice it. Imagine a school without a timetable and the chaos that would follow trying by any other means to get 100 teachers synchronised with 1,200 pupils in each of 40 periods in the week to do the right one of 30 different subjects at three different levels in 45 different classrooms. Yet the miracle of organisation that is the school timetable does this and is taken totally for granted (apart of course from by the individuals who slave through the summer holidays to put them together). And it is the same for most coordination mechanisms. We don't notice them when they work and we don't always recognise the need for them even when they aren't present and we are frantically trying to solve the problems the lack of them has caused.

The school without a timetable may seem like a fanciful example, but perhaps no more fanciful than the bank that didn't coordinate training between branches. They sent all its customer service staff off on a 3-day customer care course at the same time so there was nobody actually left to do any customer care. Failed or missing schedules are very common. Production scheduling is one of the most common areas of failure, and particularly the need to keep different operations "balanced". When this fails, we get overproduction in some areas and underproduction in others and work-in-progress piling up in factories.

As well as scheduling, production or otherwise, typical coordination mechanisms include: protocols, mutual adjustment, boundary agreements, common standards, common language, and culture.

In a training department, individual trainers decided what courses they though were needed, and then designed, set up and ran courses. Trouble was that they actually needed the cooperation of their fellow members of the training team to deliver the

courses and they needed access to shared training resources such as the training suite and facilities. Because they operated independently without coordination, facilities and people would be double booked. Each time was treated as a new occurrence, with arguments and appeals to the head of training to sort out the mess. In this case there was clearly a need for some sort of schedule for use of shared facilities but also for some protocols for negotiating and contracting colleagues to work on one another's projects.

Boundary issues are a frequent coordination issue. One of the areas where this is most prevalent is in sales territories. Where the boundaries are geographic, this is fairly easy to define, but where the boundaries are more nebulous, it is obviously harder.

An IT company increased the autonomy of its operational units but failed to put in adequate coordination. A salesman turning up at a client to sell a document management package could find that two competitive offerings from the same company had already been offered to the client. With no coordination, the company was competing against itself and wasting resources duplicating development, sales and support.

In a hospital, there was no coordination mechanism for handling the boundary between cardiac surgery and cardiac medicine. If a patient got referred to a cardiac surgeon, then they invariably got sent for surgery. Occasionally the cardiac medics would refer patients for surgery, but generally, they prescribed drugs. Patients presenting with heart problems could end up in either surgery or medicine. The basis on which this life critical decision was taken was the length of the waiting list for the surgeons. If there was a gap in a surgeon's waiting list, then the next patient would be sent in that direction. Coordination problems can have serious and sometimes bizarre repercussions.

In 1999, NASA had the embarrassing and expensive experience of crashing a probe into Mars. It emerged that the problem had been that two teams were using different measurement systems, one metric and one imperial. The thrust applied by rockets to control the probe's position for entering Mars atmosphere was calculated by one group in Newtons and by the other group in pounds force. Each assumed they were both using the same common measurement standard, but they weren't and because there are just over 4 lb force to the Newton, the probe wasn't where it should have been. The issue of coordination by common standards or rather a lack of them is very common and isn't limited to tangible things like measurement standards. Within one single company of just 60 staff, the five operational departments each used different standards for management accounts. With no common basis for comparison, it was impossible to establish which operations were actually profitable. This generated a series of ill-judged investment decisions that destabilised operations when some departments were under-funded whilst others were over-funded. Inevitably, it also created political turmoil.

Working on restructuring a bank in a post-communist eastern European country, a team of western consultants were disconcerted part way through the project to discover that whenever they'd talked with the bank's management about "cash", they had been talking about completely different things. To the westerners, cash was actual tangible money. To the eastern bankers, it was any money that wasn't part of the government's planned economy.

This sort of problem over common language isn't limited to national differences. Amongst a group of pharmacists operating within the same hospital, there wasn't common use of language – not even of their specialised technical language. Different individuals used a range of different terms for the same thing and used the same technical term to mean different things. This is a little disconcerting when we are talking about a group of people trained to be precise and scientific, all working in the same discipline in the same organisation, and especially when they are dealing with potentially life critical treatments. Problems over common language are even more common between departments and different technical disciplines and extend to the choices we make about using the same IT platforms and programmes and of course to the mental models we use. Wherever a message crosses a system boundary: between two individuals, two departments or two companies, it undergoes "transduction" a process of translation in which it inevitably gets changed to some extent. The distortion can be trivial or critical, but the purpose of creating common languages is to build effective transducers that reduce distortion as far as possible.

3.2.3.3 System 2 and the Design of Structure

Within many organisations there is a constant battle going on between support functions trying to get operations to adopt common languages and standards and operations seeking to go their own way. This is one facet of the autonomy – cohesion dilemma. Finance wants everyone to do their budgets and reporting in the same way. IT departments want everyone to use the same programs so support is easier, whilst operational departments often find reasons why they need a non-standard IT program. Both sides of this tension can be legitimate, although it's hard to see the validity of having 400 different knowledge management systems within the same organisation, as one high tech company did. Especially since the purpose of knowledge management is to allow knowledge sharing and this is prevented by system fragmentation.

Sometimes this tension which manifests as a sort of guerrilla warfare over System 2 coordination mechanisms is actually a sublimation of the autonomy – cohesion tension at the level of strategy. Operational departments denied autonomy in the direction of their operations, sometimes exercise autonomy in subverting the common standards that IT, finance, or other departments seek to impose. Whatever the politics, coordination is explicitly about restricting complexity and autonomy. The trick is to identify where there is unnecessary complexity that is destabilising operations and remove that whilst leaving differences that reflect genuine differences between operations. The payback for operational managers of accepting the reduction in their autonomy represented by coordination mechanisms is a reduction in disturbance to their operation by other departments, less conflict and much less fire fighting.

As well as being significant in their own right, System 2 mechanisms are also important in helping to work out the optimal solution to the question of how to organise the structure of value creation. Wherever possible, the basic structure of

the organisation should be worked out to reduce System 2 issues. In my view the loading on System 2 is one of the most critical design features and is probably the single most important factor in deciding between structural options.

One of the reasons that cells are so much more productive than functional layouts in engineering is because the structure eases System 2 coordination issues between operations.

3.2.4 Managing Delivery: System 3

3.2.4.1 Line Management

The structure of value creation breaks the organisation down, operational level by operational level and provides the basic seed structure for the viable system (Hoverstadt 2008; Beer 1985). The essential function of line management is to build these component operations back up into a cohesive coherent organisation that can create synergy. I use the term line management in its traditional sense, management responsible for a set of operations – sometimes the term is used to describe someone who has a personnel management role over an individual. In essence, line management is a relationship between an individual, or a team, department or division and the organisation of which it is a part, in which an agreement is made that the organisation will provide X resources in return for the individual, team or department delivering Y performance. This basic equation of resources for performance is key.

The basic design concept is extremely simple, but conventional practice goes against it in several ways, some of them fairly obvious, some of them quite subtle and mistakes in designing a structure to deliver synergistic performance are more than common.

For each set of operational activities identified in the basic operational structure, there needs to be a corresponding set of management activities, starting with the line management role to build cohesion. The purpose of this is to take a set of operations and to create synergy from them. Each level of the organisation delivers some aspect of performance that its individual components can't provide on their own. To do this, management has to ensure that when the performance of the operational sub-systems it manages are combined, they will deliver the performance this level of the organisation itself is responsible for (Fig. 3.7).

The twin strands involved in managing this relationship are resources and performance. The combination of the two into a negotiated agreement between, say, a departmental management and its constituent sub-systems is critical.

For this to work, what is needed is agreement rather than imposition. Arbitrarily imposing performance targets or budgets risks loading impossible burdens onto operations and also risks management basing their decision making and strategy on levels of performance that are not achieved and which may have been totally unrealistic.

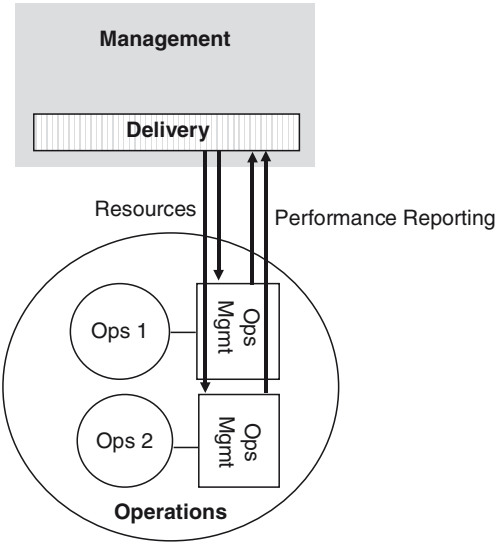

Fig. 3.7 System 3 delivery – managing delivery and Synergy, bargaining resources for performance

The traditional approach to managing resources is the annual budgeting cycle. So prevalent is this approach that it may come as a surprise to many that it is relatively new and grew to popularity in the post war drive for strategic planning. Relying on a plan meant that both performance and the resources that drive performance also had to be planned. This planning and budgeting system has become a monster that has taken over much of managers' lives – typically between 20% and 30% of senior manager's time. As well as consuming valuable management time, it encourages all sorts of dysfunctional behaviour, particularly gaming and "creative" accounting in resource negotiations. The alternative model being developed by the members of the Beyond Budgeting Round Table is one that will be familiar to many smaller entrepreneurial businesses (Hope and Fraser 2003). Typically, it uses a much more flexible approach to decide on and manage resource deployment –a combination of discretionary agreements that allow managers more autonomy within agreed limits and with the option to decide on new resourcing commitments whenever circumstances demand rather than being locked into a fixed planning cycle. This allows managers more autonomy to manage their resources flexibly whilst still leaving them accountable for results and also allows the organisation to respond quickly to any opportunities or threats that may emerge in their environment.

The key to understanding the autonomy within the recursive structure is the realisation that the management at any one level manages a set of subsystems which (a) operate within the agreed operational framework established and maintained (System 2), and (b) operate to the resource bargain agreed (System 3) beyond that the subsystems have autonomy in the way they achieve their purposes. By these means the operational variety is absorbed at each recursive level in a combination of management and self-management.

3.2.4.2 Common Failures in the Performance Management Structure

Organisations that do not measure performance do not and cannot know how well they are doing whatever it is that they do, so performance measurement has an absolutely key role to play in building an effective organisation. Which is easy to say, but is very often not done well.

The first structural element to getting it right is to not miss out performance measures. In a lot of organisations, performance measures are generalised and are not designed as specific links between elements of the organisation. For every operational element at every level of recursion, there needs to be adequate and appropriate performance measures. This means performance measures specific to each level. Performance measures follow and define the line management structure. They are one of the fundamental links between a set of operations and management at the next level. Missing out levels undermines the viability of the organisation. Missing performance measurement from one or more levels means that there are managers at that level who do not know how well the operations they are supposed to be managing are doing. This is pretty fundamental to doing the job of a line manager which is to take the resources provided by the organisation and use these to deliver performance.

The second common problem is to split resourcing decisions from performance measurement. Often operations find themselves negotiating the performance levels they are supposed to deliver, quite separately from the negotiation about the resources necessary to achieve that level of performance. Since resources come in many forms – people, skills, infrastructure, equipment, IT, money etc. all of which may be managed quite separately, combining all these resources together with an agreement about performance is not always a simple task. Where this fails, then one of two outcomes is likely, either over-resourcing – certain activities have more resource than they can use effectively, or under-resourcing – which leads to a failure to deliver performance, or at least considerable stress in trying.

Problems of misattribution are extremely common in organisations. Measures that are actually about the process carried out by department "A" are attributed to department "B". Although this may sound unlikely and the sort of thing that should be easy to spot, it is actually endemic. The reason is that predominantly, performance measurement systems are not built as feedback systems to inform decision making about specific processes and specific units. The traditional model increases the probability of sloppy attribution, because hierarchical structural models give little clue as to where processes sit and who is actually responsible for which aspects of performance.

In a small national supermarket chain, the performance of both stores and their managers were measured by sales. This is a common conflation between measuring an area of activity or process and measuring the management of that process. In this case, the measure was intended to inform the board about the performance of the store managers so that the board could take decisions about both them and their stores: which managers to promote, or fire, and which stores to expand, change or close. In reality, a store manager's area of discretion had very little impact on sales. Overwhelmingly, the decisions that did affect this measured output were taken by buyers and marketers at central office. What store managers could actually decide

about were issues around managing their staff. They couldn't decide what was sold in their store, or for what price, or how it was presented, or where it would sit in the store or when to run promotions, or any of the things that have the biggest impact on sales. These key issues were all taken centrally. So, the performance measures attributed to store managers were actually measures of central staff functions As a result, there were critical control deficits at two levels: at the level of the store and at the level of the central marketing and purchasing functions. At both these levels, appropriate performance measures were not being used to inform management decision making. The store manager's actual performance wasn't being measured, but the board thought it was and made judgements accordingly. At the same time, the set of measures that actually measured the central staff functions weren't used in taking decisions about them. Using a systemic model allows us to look at the systemic consequences of this sort of failure, and in particular what decision processes and hence what decisions are undermined by a lack of information, or misinformation. In this case, that was a whole series of judgements and decisions about individual managers, their stores, and about the management and effectiveness of a set of central functions such as buying, marketing and product positioning. In addition to the diagnostic advantage, this modelling also provides a design template for the design of more appropriate performance measures that do actually provide information where it is needed about the activities that are supposed to be being measured.

The most common problem though is the "control dilemma" (Espejo et al. 1996). Usually regarded by those experiencing it as a personality issue, it is also a structural problem and the structural solution lies in getting the structure of performance management right and specifically in monitoring. The control dilemma occurs when management worries about its loss of control over operations and so burdens operational staff with more and more demands for performance reporting. The increase in demands for performance reports is usually driven by a lack of trust that the information being given is providing either a complete picture or indeed is giving managers the answer they want. The solution is not simply to ask for more reports and more detailed or frequent performance reports, but to monitor. There is a clear distinction between 'performance measuring' and 'monitoring'.

The word "monitoring" is fairly loosely used in management. Here, I am using it to describe a particular set of activities conducted in a particular way. It is an in depth, occasional check by management, not of what their immediate subordinates are doing, but of the reality of their operations. Where performance reporting is by its very nature largely quantitative, monitoring is largely qualitative. A performance report may tell you that late deliveries go up at the end of the month. What monitoring does is let the manager who gets those reports every month, experience the semi-chaos of the shopfloor on the last Friday of the month as production tries to juggle a deluge of increasingly fractious customers and managers demanding that their job be prioritised before the weekend. Armed with that experience, the reports take on a completely different meaning. What seemed perverse and frustrating behaviour by your operations team that prevented you from hitting your target and keeping your promise to your boss is now seen for what it is, a hopeless task in the face of impossible pressures (Fig. 3.8).

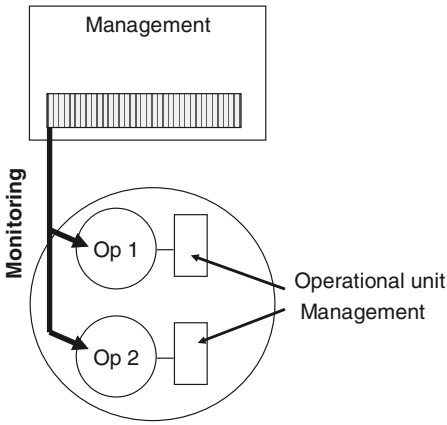

Fig. 3.8 Monitoring channel – supplements performance reporting with sporadic in-depth check of operations at next level down, bypassing one level of management

The requirements of good monitoring can be summed up in four simple rules:

1. It needs to be sporadic.
2. It needs to be unannounced.
3. It needs to skip a level of management.
4. It needs to be in depth.

Monitoring needs to be sporadic if it isn't to become too heavy handed and leave staff feeling as if they are being constantly watched. It needs to be unannounced if it is to show reality, if its predictable, then "window dressing" can hide what's really going on and the exercise becomes destructive. It needs to bypass a level of management if it is to reassure both staff and managers that management has a realistic view of what is going on in the organisation. Whole organisations have collapsed because this simple rule was ignored and managers thought that it was more comfortable just to rely on reports without checking out the reality. If monitoring doesn't jump a level of management, it provides a cover that allows unscrupulous managers to engage in all sorts of unsavoury practices from bullying, through financial irregularity to major undeclared changes in objectives, strategy or working practices.

3.2.5 System 4 – Outside and the Future – Managing Development

3.2.5.1 Systemic Function

The systemic role of System 4, the development sub-system is to ensure that the organisation maintains a healthy fit with its environment (Hoverstadt 2008; Beer 1985). In other words it has to ensure that the organisation is doing the right things

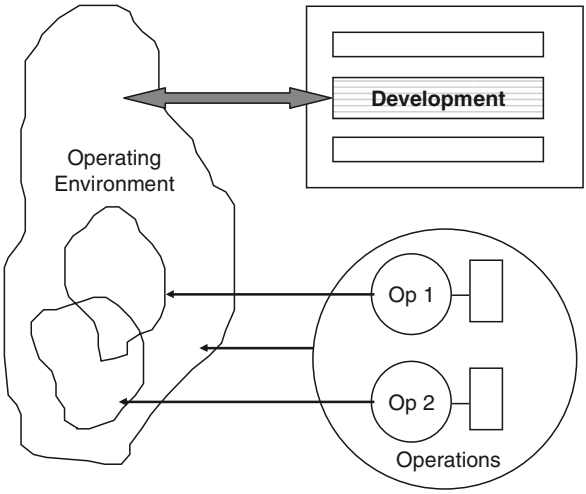

Fig. 3.9 System 4 development – surveys operating environment: Technical, competitive and market developments predicts, plans, creates the future

and able to maintain some sort of value exchange with its environment so that it can remain viable into the future. Essentially, this involves preparing the organisation to deal with changes in the environment and preparing the environment for changes in the organisation, so, predicting and creating the future (Fig. 3.9).

To do this, it has to fulfil several connected but subtly different roles, all of which relate to understanding: the future, the environment outside the organisation and the fit between organisation and environment. Typically these involve:

1. Scanning the external environment for changes or potential future changes and specifically scanning for strategic risks
2. External communications (other than those directly related to operations)
3. Innovation
4. Managing change
5. Building and holding the organisation's model of itself

Failure or weakness in System 4 is endemic in organisations and is one of the commonest pathologies encountered in looking at organisations. This is true for commercial organisations, public sector organisations and also for the third sector. To some extent, this bias against dealing with the future has been enshrined in conventional management doctrine with its emphasis on efficiency.

The failure rates of commercial organisations are evidence of just how common this systemic problem is. Of the original S&P 500 index (the Standard & Poor index of the top 500 US companies) 85% were no longer in business in 2007, so they had failed to survive 40 years. The median life expectancy of European companies used to be 60 years, it has fallen to 12.5 and is still going down. Commercial companies are failing to adapt to their environments at an increasing rate. In the public sector of course, organisational failure rarely leads to bankruptcy.

Where the System 4 sub-system is weak, disconnected or missing entirely, there are problems, both passive and active. Passively, organisations are unable to anticipate changes coming from their environment, are surprised when they happen, are unable to adapt to these changes and fail. In failing to be active, organisations don't innovate, fail to create changes in themselves or in their environment and their relationship with their environment atrophies. Where the capacity to adapt is missing or weak, systems tend to fail when their circumstances change. Of course the more complex the system the more there is that may need to adapt. So large complex systems can collapse spectacularly quickly, although it's usually easy to spot their blind spots, or lack of capacity for adaptation and consequent vulnerability, well in advance

Where System 4 fails, there are a set of mostly very common symptoms.

- Creating new products with no markets for them
- Creating markets without products to fit them
- Failing to adapt to changing markets
- Failing to adapt to changing technology
- Persisting with outdated products
- Overcome by Strategic Risks

Although, these are couched in commercial terms, exactly the same types of failure apply to the public and third sectors, so for example, an international development charity that created a type of intervention that could not be "sold" to intended users in the theatre of operations and simultaneously was failing to recognize or address new crisis areas that had opened up. International development agencies are often unable to match their response times to the pace of unfolding events – a simple lack of requisite variety.

Where development management fails, the organization fails. Often this can be seen long in advance of the actual failure and can be addressed, but many organisations with a lack of development capability are disasters waiting to happen.

3.2.5.2 Managing Change

The statistic generally quoted is that around 80% of change projects fail and when I ask groups of change agents about this, they generally agree that this figure is about right. Clearly, there is something fundamentally wrong with an approach that fails most of the time – if my car didn't get me to my destination 80% of the time, I'd think seriously about an alternative….

The traditional model for managing change not surprisingly is based on hierarchy: change is planned and implemented from the top across the whole of the organisation and cascaded down through successive tiers of management.

It is my experience that traditional change programmes ignore the essential differences between departments or teams and treat the whole organisation as if it was the same. But of course, organisations are not the same throughout. In any change programme, there are always parts of the organisation that can change more easily

than others and parts where there is a greater impetus for change. These natural differences mean that change programmes always fragment and this causes two problems. First is the perception that there is resistance. Second is the problem of consistency across the boundaries between parts of the organisation that have changed and those that haven't. These boundary issues become the grit in the change process that creates friction and drives resistance. Gaps quickly develop between teams and departments that are engaging with change and those that are not. These gaps fragment change programmes. Invariably at this stage of programme failure, managers responsible for change switch their efforts away from those areas where they are failing into the areas where they are enjoying relative success. As well as being a pragmatic response to a difficult situation, and a sensible use of their inadequate resources, this is also a very human response. Faced with a task with a high failure rate and given the option between nurturing those parts of the programme that show some hope and those where change is proving difficult, it is entirely natural to support the successes rather than confront the failures. The effect of this focus on the easy targets is to amplify the differentials that emerge, thereby further fragmenting the homogeneous nature of the programme.

In effect, change programmes that are intended to be homogeneous and 'whole company' programmes break down into discrete patches of change. This concentrates the impetus and resources for change coming from management on to just a few individuals or teams. Homogeneous undifferentiated change becomes in practice, heterogeneous, highly differentiated and discrete change. This is an entirely natural process and seems almost inevitable given the circumstances.

This isn't primarily a problem of leadership and it isn't primarily or initially a problem of resistance. In practical terms, problems occur where processes cross organisational boundaries. Where a process crosses two departments, so department 'A' hands information or components over to 'B' and where 'A' is trying to change to the new way of working and 'B' isn't, then managers are faced with a dilemma. If they carry on with change, the process will fail and if they stick with the existing process, then the change will be reversed or stalled. Faced with the dilemma – carry on with change and break a critical work process or forget the change and carry on with business as usual – managers generally take the only possible decision – go back to business as usual.

In Viable Systems terms of course, this problem of resolving cross boundary conflicts is a failure to manage the co-ordination issues and since coordination is one of the three most common pathological archetypes, it isn't very surprising that this happens.

But of course for the plan to succeed, all parts of the organisation would have to move all together and at the same pace. This is clearly unrealistic. Each department is different. Its operational demands and constraints are different. Its people are different. Their ability to handle change is different and the number of changing processes they may be handling at any one time is also different. So of course it is absolutely inevitable that they will change at different rates. Once again, Ashby's law applies, any plan that assumes change will be uniform, lacks requisite variety.

A VSM based approach to organisational transformation approaches the problem in quite a different way. It involves breaking change down into discrete, "do-able" packets

and introducing these in a sequence of planned initiatives. These allow managers to concentrate on changing elements of the organisation in a discrete way, whilst managing the interfaces between that element and the rest of the organisation so that change is not prevented by resistance through boundary issues. The sequence of change needs to be planned so that each stage helps prepare for subsequent changes either by creating structural redundancy (often in the form of management time released from fire-fighting) or by removing structural obstacles to subsequent changes.

The two key elements in this "mosaic" approach to system transformation are utilising structural redundancy and discrete packets of change. Change a component, and any other components it directly interfaces with, don't change everything at once.

Structural redundancy is about having spare capacity in the system and the amount and rate of systemic change is directly related to structural redundancy. Change requires requisite variety in the form of spare resources. The scale of change will depend on the availability of resources. Release more, and you can change more. Tackle too much and the resource will be spread too thin and nothing will work. This is of course pure Ashby's Law.

Planning mosaic change starts with a Viable Systems analysis of the organisation, both in its current state and its desired future state. Mapping these two organisational models against one another gives you a list of those bits of the organisation that will be directly affected – in other words all the parts that need to change. As well as giving you this list of potential change packages, it should also tell you about all boundary issues involved in carrying out change. There are direct transfers, such as department "A" being upstream in the same process as department "B", so if you change "A" you know "B" may well be affected. In addition, the VSM should give you all the known connections that department has with others. Existing or future co-ordination issues are particularly sensitive and important.

Following a systemic overview, the next stage is deciding where to start change. This can involve several factors. The general rule is that change must be practicable and worthwhile. Assessing practicability should include evaluating the relative capacity for change of the units concerned. The factors that affect this include:

- Group cohesion
- Experience of and attitude to change
- Skill at changing
- Quality of leadership
- Number and severity of probable boundary problems
- Management resources available to assist change

Assessment of which changes are most worthwhile at any point in the process must take into account both the intrinsic value of the change – i.e. how far it takes the organisation towards the intended destination – and critically, the capacity of the change to create structural redundancy or other factors to aid subsequent stages of mosaic transformation. The factors that aid further development will include removal of structural or process obstacles to subsequent change.

Weighing up these various factors presents quite a complex decision. In many cases, there will be an option between an initiative that is more easily achievable,

but less desirable, and one that is more difficult, but will yield bigger dividends. Although, in many cases this will be a matter of judgement, there are some hard rules that will need to be obeyed. First, the proposed change must be matched by the resources available. Although this "mosaic" inherently reduces the probability of management overstretch, it doesn't eliminate it. In some organisations, management resources available to effect change are so stretched that only the smallest systemic changes are practical. Second, there is often in major systemic change a natural chain of progress. This is almost a critical path within the plan of change, such that 'A' has to be changed before 'B' becomes practicable. This interdependency of issues or problems is a systemic feature, and is one reason for the need for a systemic overview of the organisation. Once these two basic rules have been applied, the major consideration is the creation of structural redundancy, since this can be used to create the momentum for further change.

Once change is being undertaken, boundary problems can become as big an issue as the change itself. As well as a functional analysis to identify where these are likely to occur, consideration also needs to be given to non-functional relations, and in particular the political dimension of the context needs to be considered. By making change incremental and planned, a mosaic approach helps to concentrate change management resources, so these can be more tightly focused on the interfaces of the change area and manage boundary disputes.

3.2.6 Strategic Balance

3.2.6.1 The Traditional Strategy Model

The traditional model for strategy development has three principal features. It is linear, it is deterministic and it is based on a hierarchical model. In other words, a management team or board decide a fixed goal or vision and set down a straight path of things the organisation needs to do to move towards this fixed goal and hopefully arrive at the desired destination.

This linear deterministic approach to strategy has been the prevailing paradigm since the 1960s and is based on an assumption that management can reasonably decide on a set of goals about the future of the organisation and that performance can then be measured relative to these goals. This deterministic approach is usually encapsulated in some sort of methodology that follows a linear path that runs: vision, mission, strategy, targets, performance measures. Each step is determined by reference to the previous step (Fig. 3.10).

Fig. 3.10 Traditional model – strategy and performance management

The three basic elements, determinism, linearity and hierarchy are mutually supporting and consistent. If you can determine a goal – your vision – into the future then logically the rest of the strategic process should be a linear development that follows from that vision to get you there. So determinism requires linearity and of course, linearity requires determinism. You can't have a linear process unless you know the destination. Similarly, hierarchy supports them both. It's difficult to get a large group to agree on a single vision. For that, you need a small group of decision makers or even a single (preferably inspirational) leader. Once fixed, the rest of the organisation is targeted by the hierarchy to meet the vision. It is difficult (but not impossible) to do deterministic strategy without a hierarchy and its difficult for a hierarchy to do strategy in any other way.

Unfortunately though, this traditional model has several very major shortcomings. The most important of which is that it very rarely works. Figures vary, but most surveys conclude that over 90% of strategic plans are never implemented and one survey found that 98% of strategic plans were not carried out.

The roots of failure of the traditional model are found in its features, its linearity, its determinism and hierarchy. Firstly, the deterministic approach assumes a degree of environmental stability that is rarely found today. Following a goal that was set in a strategy formulated often years earlier is only sensible if the world still looks the same as it did when the strategy was decided. In many environments, in both the public and the private sector, this is rarely the case. If our operating environment changes faster than we can achieve our strategy, then that goal based strategy is likely to be irrelevant and can even deliver us prepared for a world that no longer exists.

Take for example a leading electronics firm specialising in defence systems. They followed a goal centred strategy to become a global player in the communications market and invested heavily in a market that was new to it – optical fibre technology. By the time the strategy was fulfilled, the market for optical fibre had already peaked and the company was poised for a world that no longer existed. Typically, the planning cycle is run on an annual basis, which means that there can be a very long time lag in the feedback process that tells you that the plan isn't working. In the case of the electronics company, the strategy proved fatal.

Secondly, because it is a linear model, it has performance measures as an output of strategy. You set measures that tell you whether your strategy is working. So this approach ignores the need for performance measures to inform the strategic process as an input. In the absence of suitable performance information, strategy is inevitably misinformed and the result is a proportion of strategic plans that the organisation does not have the capability to deliver. Performance measures need to be, not merely an input to strategy rather than an output of the strategic process, but also designed specifically to provide the information that strategic decision making will need. So a strategic process that reduces performance measures to being an output has problems.

The problems with the linear deterministic model are compounded because of its connection with the hierarchical model of organisation. This may seem paradoxical because the whole point of a hierarchy is to centralise decision making,

precisely to make it easier to set strategy. Hierarchies are designed specifically to be unstable structures that allow a single individual or small team to move a whole organisation. So, it may seem odd that in practice, they aren't actually very good at formulating strategies that actually work. What hierarchies are really good at is taking decisions. What they aren't good at is taking decisions that actually get implemented. There is a strong inverse correlation between involvement in a decision process and rejection of the decision or resistance to it. The more hierarchical the decision process, the fewer people involved. The fewer people involved, the less the rest of the organisation will trust it. The less they trust it, the less likely they will be to carry it out and actually implement it.

3.2.6.2 Strategy as an Emergent Property of Structure

The connection between strategy and structure is both complex, and dynamic. Strategy often determines structure, and drives changes to the organisational structure. Less obviously, organisational structure also has an enormous effect on strategy. The two are linked together not just at any one point in time, but also through the passage of time, and this is the dynamic that drives the evolution of organisations.

It is easy to see how strategy drives organisational change. The outcome of strategy is often either a new direction for the organisation or a change of pace. To put these into effect requires some changes either to formal structures (departments, teams etc.) or at least to work patterns and communications.

What is often unseen is the way in which structure determines strategy. The strategic options open to an organisation are limited by the information that is fed into the strategic decision making process. These limitations are not arbitrary, they are structural. Messages come in to the organisation from its environment all the time, some good, some bad, but the organisation can only hear the sorts of messages it is structured to hear. If there isn't a part of the organisation that is tasked with hearing messages on a particular set of topics, then the organisation will not hear those messages. Individuals within the organisation may hear them, but the organisation cannot, unless it has structured itself to hear them. The information may come in to an individual in the organisation but then it just dissipates through the organisation because there is nowhere for it to go.

This may seem bizarre, but we experience it on a regular basis. Ever tried complaining to organisation that doesn't have a customer complaints department? Gradually it dawns on you that you are engaged in a totally futile exercise. As you try to explain to someone in the organisation what has gone wrong, they wait for you to get off the phone so they can get on with their job – which isn't dealing with your problem. The poor employee has heard your problem, but without a structure, it rarely gets any further because they have nowhere to send the information. Exactly the same principles apply to the classic strategic topics. Without some part of the organisation tasked with understanding the market, or changes in technology or economic trends or competitive pressures, decisions will be taken in absolute or relative ignorance of those key topics.

So, the classical assumption that the organisation is an outcome of the strategy because you change the organisation to suit the strategy is true, but it's only half-true. The reverse is also true, because the organisation's strategy is also an outcome of its structure. This may seem contradictory but of course, what it means is the structure and strategy are linked together in an evolutionary cycle in which the structure affects the strategy, which affects the structure. The obvious outcome is that the organisations follow evolutionary pathways that are largely determined by who they are now and they progressively evolve to become more "themselves". This is a natural evolutionary process of organisations structurally coupling to their environments and is quite different to trying to force change through goal setting. This approach does carry a risk of the organisation becoming increasingly culturally and informationally closed. The way to avoid this danger is by ensuring that the intelligence function is operating effectively. If it is pulling in diverse information about what is happening in the environment and specifically monitoring strategic risks then threats to the relationship that is emerging with its chosen environment can be avoided.

3.2.6.3 Strategic Conversations

Systemically, good strategic decision making relies on balancing the capabilities of the organisation as it is now, in its current operating environment, against the demands that it needs to address in its environment and in the future. As the environment changes, as demands change, those changes need to be detected, or better still anticipated, and brought into the strategic debate. Seeing a need for change creates a "strategic gap", a gap between what can currently do, and what we have identified that we are going to need to be able to do in the future. The process of strategic decision making is then to work out which of the identified strategic gaps the organisation should close, and how this should be done. And this is what strategic decision making does, it opens and closes the strategic gap to drive the organisation's continuous evolution and adaptation through time. Closing the gap is primarily the job of "delivery management" (System 3). Opening up the gap, in the sense of perceiving it, making it explicit to the organisation, and making practicable is the job of "development management" (System 4).

All management disciplines tend to have their own areas of interest and their own language. Consequently communication between them can be difficult. Marketing and operations don't talk the same language. They don't see the world in the same way and indeed aren't even looking at the same bits of the world. Both are different from the finance department who speak another language and view another landscape. And yet, despite these very real differences, we need all these different specialist interests and others to come together if we are to come up with strategies that are practicable and appropriate. Robust decision-making, coming up with a strategy that actually gets implemented, requires that all aspects of the strategy are examined (Fig. 3.11).

If our strategy involves introducing a new product currently in R&D for example, then R&D (System 4) need to check with operations (System 3) that they can produce it. Finance needs to be involved over both short term cashflow implications

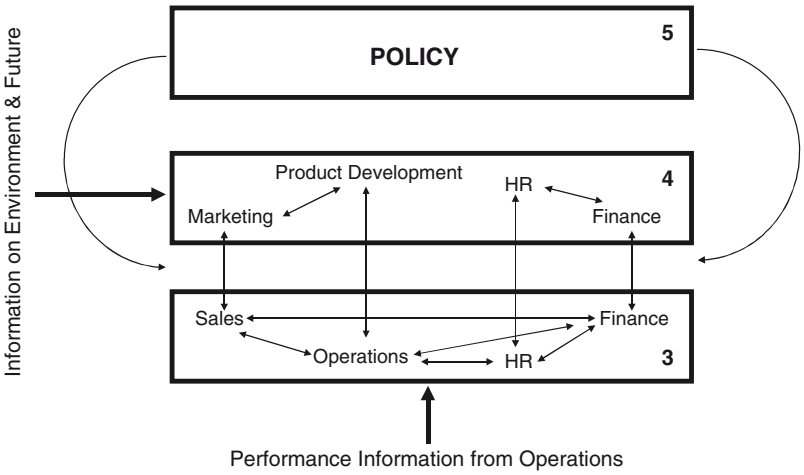

Fig. 3.11 Management decision structure. Typical set of strategy conversations, connecting different disciplines on a range of interdependent issues. With System 5 ensuring the decision structure's integrity by integrating internal and external issues

(System 3) and longer term investment planning (System 4). HR may well need to be involved on short term staffing arrangements for operations (System 3) and for either recruitment or training (System 4) if current capability doesn't already exist. Similarly marketing (System 4) needs consulting early on about market opportunities for the new product and sales (System 3) about how the new product might disrupt existing sales. This is a complex set of conversations and not one that can follow any pre-set process. In reality, these are interdependent not independent issues and the outcome of one conversation may require us to go back and revisit another. So operations may agree with HR that they need to recruit new staff and retrain others to make the new product, but a conversation with finance may force them both to think again.

3.2.6.4 Getting the Balance

Within this decision structure there are two very different types of management behaviour. On the one hand, there is that part of management engaged in running the organisation as it is now, and all the behaviours that go with that. Trying to optimise, measuring performance and resource usage and seeking greater efficiency. On the other hand, there is that part of management engaged in developing the organisation and creating the future. This involves scanning trends, analysing market needs, creating options, researching and developing products, technologies and markets and planning how to move forward into the future. These two sets of activity are both essential, but are pulling in opposite directions.

Good management of the organisation as it is now, so it operates efficiently is essential. Without it, the organisation will struggle to survive in the short term and as the saying goes "without a short term, there is no long term". Equally important however is the need to envisage and create a future for the organisation. Without development and adaptation, the organisation will lose its "fit" with its environment. It, will fail to provide goods or services that a changing world values and will die. So both types of management activity are essential to survival.

The problem is that whilst both are essential, they are also in tension, and in several different ways. They require different types of thinking, so individual managers are predisposed to one or the other. They require different types of information, so the organisation's management information system can provide biased support. Most obviously, they are in tension because creating the future inevitably involves reducing the efficiency of the present.

Any action to change the organisation to meet future needs, inevitably involves using resource. No adaptation – not even one aimed at improving efficiency is entirely cost free. There always has to be at least some "pump priming" and if we are talking about major strategic change in the direction or identity of the organisation, then this generally requires a significant call on resources and particularly management resource. All of this inevitably reduces the short term efficiency of the organisation and disrupts managers' attempts to run a "well oiled machine". As well as diverting resource, it also diverts attention away from the efficiency issue and so has a political effect of tending to undermine the importance of those managers focused on the here and now.

The result of this dynamic tension between stasis and change is that organisations are often unbalanced to either one side or the other. This strategic balance is acted out in the 3–4–5 balance. Where this favours System 3, strategic decisions tend to be of the "do more of the same" or "do less of the same" variety, so strategies are about growth or retrenchment, but not about doing something different. This can be fatal if the nature of the environment has changed and requires something different from the organisation. Where the imbalance is in favour of System 4, then organisations can develop strategies that involve changes of direction, the development of new markets, new innovations, or new technologies. This can be fatal if the new direction isn't within the capability of the organisation and the strategy is unrealistic or unachievable. Where it is in balance, organisations are able to develop strategies that do involve genuine changes of direction but ones that are within their capacity for change.

3.2.7 Identity and Governance: System 5

This is a tricky section both to write and to get to grips with, because we need to address two different issues in parallel. Firstly there is the issue of "how to analyse/design System 5 of the VSM" – the bit responsible for governance and identity. Secondly there is the issue of "how do we define the identity of the system we are analysing/designing using the VSM so we know what we are looking at". This a paradox, because in the

VSM we are not just talking about modelling a system to which we have ascribed an identity, we are looking at how an organisation modelled as a system creates, maintains and recreates its own identity for itself. Inevitably then, these two: the identity ascribed by the modeller and the identity the organisation creates for itself, connect at some point and need to connect. So in this section I'll be flipping between these two perspectives. In practice, this flipping is a necessary sense check – is the model of the system we are building actually aligned with the organisation's own self construction of identity. This does not mean, modelling the system as it has been described in some sort of "mission statement". Organisational identity isn't the prerogative of senior management, so the comparison is between the view of identity the modeller started with and what the analysis reveals of how the organisation builds its identity. A further complication is that there are different ways of defining identity depending on the purpose of modelling.

3.2.7.1 Defining Identity

When building a VSM I use two different approaches to defining the identity of an organisation as a system, depending on the purpose of the modelling exercise (Hoverstadt 2008). The first is a fairly conventional one which is to define the system by "purpose", by what the system does. The second is to define it by its "structural coupling".

The first uses a formula "a system to do x by means of y for purpose z". This type of definition is embedded in a set of stakeholder relationships which can be categorised using the mnemonic TASCOI (Espejo et al. 1999)

T = Transformation, what the system changes from what into what
A = Actors, those carrying out the transformation
S = Suppliers to the transformation process
C = Customers, those in receipt of the transformed product
O = Owners, those responsible for ensuring it happens
I = Interveners, those with an interest in the process

I use the approach of defining by purpose when designing a new system and when problem solving. In this latter case, because problems are not things in the real world, they are essentially a gap between how someone thinks the world should be and how they perceive it to be, defining perception is critically important. More specifically the perception of the person who wants their problem solved as to the purpose of the system is critically important.

I use definition by structural coupling in any other situation and specifically when analysing an existing organisational system. Rather than looking at purpose and from a specific viewpoint, this is definition by relationships. So the organisation is defined as a system that has a set of relationships with different parts of its environment in which some value exchange happens which affects the structure of activities (in other words causes you to do something different). The advantage of this approach is that it involves modelling the same organisation from multiple viewpoints at the same time. The organisation will have multiple relationships each

of which may have radically different perspectives of purpose, value and significance. Since the interaction of these perspectives and value judgements may be significant, this approach gives a more rounded model than modelling from a single named perspective.

Defining identity by structural coupling is actually to define the system in a very literal sense – to define by its de-fined limits, by its boundaries. Whenever we put a boundary around something, and in our case around a part of an organisation, we are defining something. We are deliberately separating what is inside from what is outside the boundary. We are saying that inside the boundary is different in some way from everything else outside. This creates an identity for what is inside. This happens whether we like it or not, every time we build a boundary. Every time we set up a new team or department, or business unit, we create a new identity. So, identity is an aspect of structure and the boundaries we create or the ones that we or other people recognise.

The significance of boundaries and identity here is that confusion over boundaries and therefore identity is becoming an increasingly common source of mismanagement as organisations adopt new forms in an increasingly complex global environment. From a modelling point of view, when you have an organisation that has outsourced key systemic functions, there are sometimes difficult decisions to be taken as to where you define the boundary of the organisation.

3.2.7.2 Systemic Function of System 5

The systemic function of "System 5" is to do three distinct but related things. Firstly it has a governance role for the organisation, ensuring that the organisation is functioning as a system capable of managing itself and of steering a course that will keep a healthy fit between the organisation and its environment (occasionally, this will involve dissolving the organisation if that is the most appropriate thing to do). Secondly, it needs to create, maintain, or recreate the identity of the organisation. Thirdly it needs to maintain an understanding of the relationship between the system-in-focus and the meta-system, the system within which it is embedded.

Each of these roles is fairly nebulous and in practice, this is often the most difficult of the VSM sub-systems to identify, not least because when it's working well, it's nearly invisible.

Given the elusive nature of System 5, it is often difficult to spot where the capacity is that performs its roles and in practice, it's often easier to look for connections than actual tangible resources like a team. Once we get high up an organisation, we'd expect some sort of "board" which should be fulfilling some of these roles. If we look at the specific System 5 task of ensuring there is a balanced debate between System 3 and System 4, within a board (where we might expect such a debate to take place) this is the role fulfilled by a good chairman.

In formalised project management environments, "project boards" nominally at least play some of these roles, although often this is pretty nominal. Once we get down to the level of a department or team though it may be much less formalised.

If we think about the roles of System 5 in the context of a board (because that's where it's most obvious) and the connections needed, and how we might spot those, then there are three we need to look at:

1. The governance connection – specifically to maintain the balance between System 3 and System 4 in formulating strategy.
2. The governance connection into lower levels of recursion in the organisation to hear alarm calls that levels of management might filter out.
3. The connection to the wider system within which our system-in-focus is embedded.

In looking at the connection to ensure a balanced board debate to create a viable strategy, when you watch a well chaired meeting of this sort, the role of the chair is almost invisible, but everyone has had space to voice their views, and been listened to.

There has been much talk in the public sector about whistle-blowers. Much of this has been slightly schizophrenic, they are approved of by ministers when the whistle is being blown on public sector bodies behaving in a way ministers don't approve of, whilst they are pilloried when the whistle is being blown on ministers themselves. Part of the System 5 governance role is to be able to hear these messages from deep in the organisation that things are not as they should be. External whistle blowing – going public is a sure sign that this function isn't being discharged. There is no single right answer as to how this particular connection should operate, some CEO's and chairmen do it effectively by "walking the floor" by physically making themselves available to staff, but there is a limit on the size of organisation you can cover effectively in this way. Informal networks can work well and in organisations, there often are particular individuals or chains of individuals who fulfil this systemic role, conveying messages up the organisation that all is not as it seems or is reported and that there are problems being hidden.

The third connection to understand the system-in-focus's place within the wider system in which it's embedded – which of course impacts on its understanding of its own identity. Formally, this includes things like engagement in industry bodies, professional or trade associations and societies. Benchmarking exercises to compare your organisation to others in the same sector can also help. It explains the systemic role or potential systemic value of the time CEOs and chairmen spend on golf courses and other similar apparently trivial activities. This sort of networking with peers outside of the organisation's immediate stakeholder network is vital to the role.

For each of these types of connection, there are equivalents at any level of recursion. Some of these will be easier to assess (like being accessible to danger messages) others less so.

3.2.7.3 Symptoms of Failure and Pathologies

Where governance fails, we see the disintegration of the organisation. The most sensitive areas are the balance between delivery (maintaining the status quo) and development (change) and the measuring and monitoring of performance.

Where either of these are under-resourced, disconnected or simply missing, the organisation is likely to fail whenever circumstances become unfavourable. These are disasters waiting to happen. A common manifestation of this is the "Death Spiral". Systemically, the problem starts with a failure of governance to ensure that there is a balance of strategic decision making and specifically a failure to address external and future factors. When the complexity of the environment changes, this isn't noticed. Because of the failure to prepare adequately, operations respond to the environmental changes erratically. This triggers either inter-unit instability or intervention by higher management (control dilemma) or both. This reduces the ability of the organisation to respond at both the operational and the strategic level. As a result, operational responses to environmental change are inadequate and the organisation starts to fail. If management notices – and often they don't – they usually go into crisis mode – bunker mentality. This further reduces their ability to address the problems and reinforces the initial isolation from external intelligence. At this point, the organisation can usually only be saved by external intervention. Either an injection to the management team, or a further change in the environment is needed. In other words, organisations in this state only survive by luck. The process starts with a failure of governance (Fig. 3.12).

For me, the Death Spiral illustrates two things well – apart of course from showing this all too common mode of collapse for organisations. The first is just how critical Governance is for the sustainability of organisations and the second is the systemic nature of organisations and the problems they have.

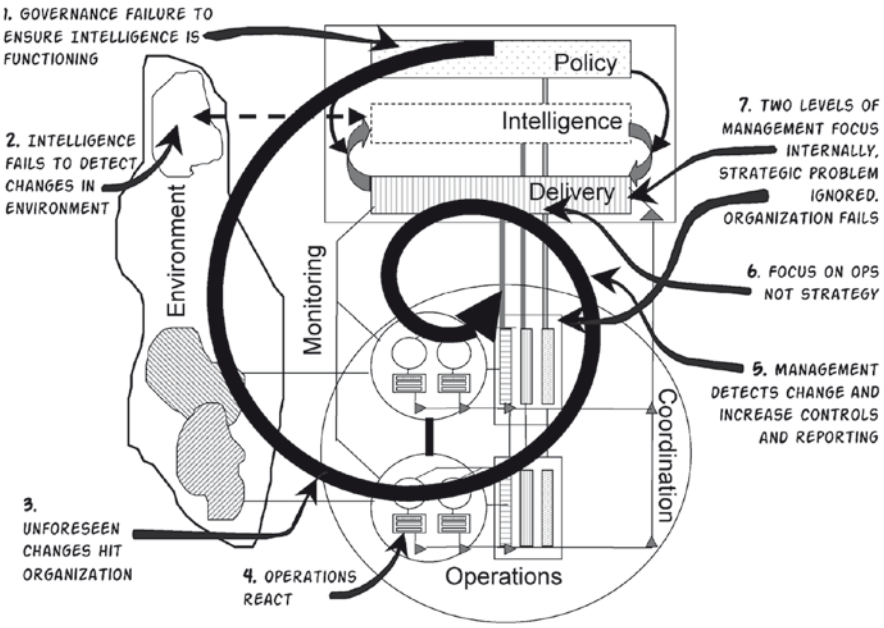

Fig. 3.12 The Death Spiral

Where governance is weak or focused on compliance or on internal control, in other words for most organisations, the death spiral is a disaster waiting to happen. As our organisations become increasingly complex internally and operate in increasingly uncertain and turbulent environments externally, failures in the critically important role of governance will continue to drive up the failure rate of organisations.

The Death Spiral is also a good example of how organisations operate and fail systemically. Failure of one part of the organisation can cause a chain reaction throughout the rest of the system. A possibly long standing failure of governance creates a flawed strategic decision structure, which goes unnoticed until the environment changes which then causes a sudden failure in operations which in turn triggers a collapse of strategic response. As the spiral winds its way inexorably inwards, so the time the organisation has to react shortens. Usually, the governance problem could have been addressed years in advance of the crisis, but by the time the crisis hits the senior management team cut off in their mental bunker, there may be only days or occasionally even hours to save the organisation.

This means that understanding these systemic linkages is critically important and also that doing so gives us the opportunity to deal with many organisational weaknesses well in advance of their manifesting as actual problems. This all relies on understanding how different aspects of decision making need to interact, understanding how these relate to operations and to the environment. In short it depends on understanding how the organisation operates as a system. In this way, the model of organisations outlined in this book is totally different to other models of organisation such as the hierarchical model which model organisations as static structures. The essence of this approach is that it models organisations as dynamic systems co-evolving with their changing environments.

3.3 Reflections

3.3.1 Model or Methodology

It's easy (particularly in this sort of context) to lose sight of the fact that the VSM itself is a conceptual model not a methodology – you do need some sort of methodology to apply it, but it is not itself a methodology, it's a model of organisation. As a model, it encapsulates some principles, laws and axioms of organisation, but it isn't itself a methodology – a way of addressing a situation, much less a method – a step by step process of investigation. The differences this makes are profound. A lot of systems knowledge is bound up in methodologies and some of the rest of this reader covers some popular methodologies, but VSM isn't really like that. It has often been criticised for not being a methodology on the grounds that this makes it harder to use. Well it does and it doesn't.

I think there are two reasons why VSM practitioners have often been reluctant to go down the methodology route. The first is that methodology tends to be

focused around one type of application – say problem solving, whereas people use the VSM in a very wide range of fields and for a very wide range of purposes, so methodology can be perceived as restricting. The second is that the basic methodology for VSM is so simple that it hardly warrants the term. Essentially VSM methodology consists of pattern matching – you take a real world situation, compare it to the VSM as a normative model, see where there are differences between what you perceive in the world and the normative model, and then see whether those differences tell you anything useful or interesting about what you see as this real world situation. Which is pretty simple, but also pretty non-specific.

Beginners often struggle with the question "so where do I start then?" and of course the one sentence methodology above doesn't help you with that. The good news is that it doesn't necessarily matter where you start although some routes to building a model will take you much longer than others. Some practitioners have resorted to some quite linear methods to help and these take you through a series of modelling steps to build up a "complete" model. Some of them are even quite good.

But, if you talk to seasoned users, they don't usually do that. They very often don't even bother building a "complete" model. They just seem to go quite quickly to the core of the organisational issue, whether this is to do with analysis or design and they focus on that. Part of the reason is that they tend not to use linear methodologies and partly it's that they often start their analysis in a completely different way to beginners.

Beginners to the VSM will often see it as a sort of static model, maybe not much more than a rather complex organisational chart, and proceed by doing a "filling in the boxes" exercise. But of course this is only part of the picture. Systems models tend to be more about the connections than the things they connect – after all, it's the way things are connected that gives rise to emergent properties. In the VSM, although it's presented as a graphical model and the connections between the component subsystems of the model are shown as lines or arrows, what each of these represents is a feedback loop and a complexity equation. What we are actually looking at is how the organisation reconciles the fact that the environment is more complex than itself and that its operations are more complex than management. Each of these complexity equations has a dynamic or expresses a dynamic tension. If we can maintain requisite variety, we have some degree of order, if we don't have requisite variety then we have chaos. In the terms used by complexity theorists, Ashby's law defines the edge of chaos and what the VSM models is the organisation's ability to walk the edge of chaos. It's very far from being a static model. So, experienced modellers become acutely sensitive to spotting chaos or stasis, either actual or potential and analysing the systemic roots of these two weaknesses. At the same time as they are "filling in the boxes" they look for these imbalances and focus their attention around those. In problem solving of course this is particularly useful, most presenting problems in organisations have a requisite variety component – either as their root cause or as a symptom, so sniffing these out and chasing them down can often provide blisteringly fast diagnosis.

So the pattern of investigation is often much more of a "natural" hunting pattern than a linear methodology. Different areas of the model may be briefly explored – maybe an initial sketching in of the boxes, and if that shows up a missing connection or a lesser failure to maintain requisite variety, then that deserves deeper attention, otherwise move onto another area of the model and sketch that out.

The dynamics of complexity imbalances also provide a rich source of prediction. If for example you know the environment is changing faster than the organisation can plan its response, it isn't hard to work out the sorts of symptoms you are likely to find and to check out whether those are present. As well as being a reassuring sanity check on your analysis for the modeller, this loop from initial analysis, through checking symptoms and feeding back into the model can have dramatic impact with anyone in the organisation. Being able to explain the systemic causes of existing pain and predict future problems can have a very strong emotional impact and help "buy in" to the model and the modelling process.

So experienced modellers tend to flip between modelling the static elements of the model (people and teams fulfilling specific systemic roles at particular levels of recursion), to the dynamics and back again. To the untrained observer this process can appear rather arbitrary. The comforting thing is that while in any particular situation there are some routes though an analysis that may be very much faster than others, the slow routes should still work, they just take longer. This means that for the inexperienced practitioner, the question of "where do I start?" is much less important than with many systems approaches.

Following a "dynamic" approach is easiest when problem solving, as very often the presenting problem will be analysable as a variety imbalance issue. Once this is established, the next step is to model which variety balance it is, so where in the VSM it fits – for example, is it between two operations or between operations and the operating environment, and then to model the static elements around that imbalance.

When tackling an analysis from the other direction, starting with the static structure, most modellers may start by defining the identity and then unfold the recursive structure of primary activities that fulfil that identity. From there the other subsystems from 2 to 5 are added. One approach is to progress from static to dynamic across the subsystems. So once the identity and recursive structure are defined, first look at whether there is capacity to fulfil the systemic role of each of the five subsystems, then whether it has the connections it needs as set down in the graphical model, then whether there is requisite variety across each of these relationships. This progressive approach has two advantages, it's thorough and it can save time by leading you quickly into an analysis of requisite variety and therefore of dynamics. Obviously, if there is no capacity to maintain a systemic link, there can't be any connection and if there isn't a connection where there needs to be one, then the system cannot have requisite variety. So this approach can lead you fairly quickly from tangible aspects of analysis – "is there anyone fulfilling this systemic role?" through to "if not then we can't have requisite variety so are these sorts of issues being dealt with?" to "so does that explain the chaos/pain/uncertainty?"

3.3.2 VSM as a Source of Methodology

VSM has been described as a "master organising idea" and what was meant by that was that in providing a model of organisation, it offers a framework to understand how other management approaches fit together (or don't). This can be massively important.

For example in a recent (2009) project on Business Intelligence in commissioning in the NHS, the distinction which is very clear in the VSM, between Business Intelligence information (System 4 to the environment) and performance management information (System 3 to System 1), was critical in the diagnosis that some trusts were missing the intelligence information altogether because they'd confused the two. With no clear model of the difference between intelligence and performance measurement, all "data" was treated as if it was the same – irrespective of its source or true meaning. What this meant was that instead of taking decisions based on System 4 intelligence about the health needs of the population and commissioning services to address those, they were taking decisions based purely on how hospitals performed in discharging their contracts. The implications of this were that they would base their decision making purely on performance management data about the status quo, leading to a cycle of repeating old patterns. With no intelligence to tell them when provision was out of step with need in the environment, the status quo could not be successfully challenged and areas of health inequality could not be addressed. In practice, this diagnosis was confirmed, trusts that didn't incorporate intelligence, but only relied on performance data were unable to take "cycle breaking" decisions.

As well as acting as a framework that provides a context for other approaches, the VSM has also been a fruitful source to develop methodology for some common management issues. So there are "VSM" derived methodologies for: software development, change management, performance measurement, strategy development, strategic risk, innovation strategy, knowledge management, finance management, management accounting, and governance. Generally these are radically different to the traditional models and methodologies which tend to be linear and deterministic rather than systemic.

3.3.2.1 Different Applications

I've talked about VSM in terms of an organisational model to look at "human activity systems" and the emphasis has been on formalised, systems that the casual observer would recognise as entities in the real world – companies, hospitals, charities that sort of thing. But of course, VSM isn't just a model of organisations it's a model of ***organisation*** and it's relevant in many other domains.

Firstly it's widely used to model systems made up of a number of organisations, from the socio economic system that is a nation state downwards. So for example one project I was involved with was to model the system for decision making about nuclear waste in an EU country. This wasn't a single "organisational entity" it was a system composed of around 20 different organisations interacting to do decision making about what to do (or indeed not to do) with nuclear waste. Whereas other organisational models really struggle with this sort of situation, the power of the VSM to make sense of big complex systems really comes into its own.

The first book Stafford Beer wrote on the VSM was "Brain of the Firm". In it he mapped the VSM to a known viable system – a human being. The VSM is a

model of the human nervous system. For many years, “Brain of the Firm” a book on management was a standard teaching text in medical schools as the best available text on the working of the human nervous system. In fact more copies were sold to medical schools than to business schools. Since then, it’s been used to model a number of biological systems, from single celled organisms to bee colonies. The “mosaic” approach to change which is based in VSM is also the way species evolve – structurally redundant components get recycled and used for new and different functions to create new capabilities.

As well as being a biological model of the individual human, VSM can also be used to understand individuals as activity systems. A whole variety of personal issues such as identity – the maintenance of structural coupling, personal integrity, “work life balance”, personal decision making, managing personal change and relationships can all be usefully modelled with VSM. I personally struggle to maintain a balance between System 3 and System 4 – I find it all too easy to create more new opportunities and developments than I can deliver on. In true Conant–Ashby style, modelling that dynamic does help me to manage it better.

The VSM has also been used to design software. There have been databases and operating systems designed using VSM as a structural model. Two of the holy grails of software development are adaptive self regulating software and reusable software. Operating systems designed using VSM made significant steps towards those goals. The database design approach that uses VSM allows for a much more evolutionary approach to handling big systems development and integration than conventional approaches. The adoption path here has been software designers and computer scientists looking for a model that would allow them to do what they needed and finding VSM rather than VSM enthusiasts running off to develop some software in a fit of enthusiasm – in other words, it has been adopted by necessity.

To my eyes the oddest application of VSM has been in the arts. Whatever you think of their music, once you know that both Brian Eno and David Bowie are VSM aficionados, its hard not see their musical careers in a completely different light. In the visual arts, there have been a number of artists who have incorporated some of the concepts underlying VSM into their work and more broadly, cybernetics have been widely taught in art colleges where it gave a handle on the issues of how identity , the mutability of identity and meaning is constructed. In popular culture, cybernetic concepts have been a major influence on science fiction, not merely at the very superficial level, but again in the issues of the construction and self construction of identity and meaning.

3.3.2.2 Ethics

It may seem odd that quite a lot has been talked about and written on the VSM and ethics. Indeed some of the early critiques of the VSM were based on an ethical criticism of what it was supposed (incorrectly) that the VSM represented – an improved model of hierarchy that would help those in power to control people more easily. Conversely, some have seen the VSM as negating all centralisation

and taken the handle it gives on the significance of autonomy as an anarchist's charter. In fact of course the VSM is not a charter for either anarchists or dictators and in both these cases, it will show the limitations of each position, although the "anarchistic" faction have been less sensitive to the messages the VSM has about the limits of anarchy than the dictators have about what it has to say about the limitations of totalitarianism. But then I guess you only get to be a dictator by being intolerant of any limit to your power, real or implied, so they would be sensitive.

There's another aspect to the ethical debate which rarely gets mentioned and that's to do with ethics and recursion.

As you go up levels of recursion, you go up levels of logical concern. If you take a particular level of recursion as your system-in-focus and define its identity in terms of "what" it does, then the level below will be defined by "how" it does it and the level above by "why". These are different levels of logic and at each level the system engages with different issues and looks at a slightly different bit of the environment. What may seem appropriate at one level may seem very inappropriate when viewed from a different level even within the same system, because another level of recursion will see different consequences of the same action. A lot of common ethical problems fit into this architecture, something which is seen as a good at one level of the system which is seen as bad at another. This isn't just a phenomenon you can observe, it provides a basis for predicting and modelling ethical issues, both their appearance and their structural and systemic drivers. This in turn feeds into the design and running of governance.

References

Ashby, W. R. (1956), *An Introduction to Cybernetics*, London: Methuen.
Beer, S. (1959), *Cybernetics and Management*, London: English Universities Press.
Beer, S. (1966), *Decision and Control*, Chichester: John Wiley.
Beer, S. (1974), D*esigning Freedom*, Chichester: John Wiley.
Beer, S. (1978), *Platform for Change*, Chichester: John Wiley.
Beer, S. (1979), *Heart of Enterprise*, Chichester: John Wiley.
Beer, S. (1981), *Brain of the Firm*, 2nd Ed, Chichester: John Wiley.
Beer, S. (1985), *Diagnosing the System for Organisations*, Chichester: John Wiley.
Beer, S. (1994), *Beyond Dispute*, Chichester: John Wiley.
Conant R. C. and Ashby W. R. (1970), *Every Good Regulator of a System Must be a Model of that System*, Int J. Systems Sci, vol. 1 No. 2 p. 89–97.
Espejo, R., Bowling, D., and Hoverstadt, P. (1999), *The Viable System Model and the Viplan Software*, in Kybernetes, Vol 28 Number 6/7, 661–678.
Espejo, R. and Harnden, R. (Eds) (1989), *The Viable System Model: Interpretations and applications of Stafford Beer's VSM*, Chichester: John Wiley.
Espejo, R., Schuhmann, W., Schwanger, M., and Bilello, U. (1996), *Organisational Transformation and Learning*, Chichester: John Wiley.
Hope, J. and Fraser, R. (2003), *Beyond Budgeting*, Boston, Mass: Harvard Business School.

Hoverstadt, P. (2008), *The Fractal Organization*, Chichester: John Wiley.
Liker, J. (2003), *The Toyota Way: 14 Management Principles from the World's Greatest Manufacturer*, New York: McGraw Hill.
Monden, Y. (1983), *The Toyota Production System*, Industrial Engineering and Management Press.
Sloan, A. P. (1962), *My Years with General Motors*, New York: Doubleday.
Taylor, F. W. (1911), *The Principles of Scientific Management*, New York: Harper and Brothers.