HSE
Health & Safety Executive

Development of an information-based approach to self-regulation of health and safety in small firms

Prepared by
John Kingston Associates
for the Health and Safety Executive

CONTRACT RESEARCH REPORT

330/2001

Development of an information-based approach to self-regulation of health and safety in small firms

Dr J Kingston-Howlett
BSc MSc(Eng) PhD MIOSH
John Kingston Associates
Taylor Cottage
76 Blymhill Lawn
Shifnal
Shropshire
TF11 8LT
United Kingdom

The purpose of this study is to consider the possible advantages of an approach to regulation that concentrates upon the information handling patterns of small firms. The general orientation accords with the Robens philosophy of self-regulation and attempts to explore the HSE's role as a designer of self-regulation in small firms. This research does not review HSE's current practices and confines itself to the value that a different set of concepts might offer to HSE's work.

This paper presents a simplified view of a cybernetic approach to regulation. Cybernetics is a branch of systems theory that is concerned with control in systems and is designed to address dynamic complexity. This considers regulation as essentially a problem of directing and processing information. Agencies such as HSE recognise the financial pressures on small firms and take account of them in judgements of regulatory impact. However, less visible are the constraints of information processing and these, though less visible than financial resources, are argued to be at least as significant. The importance of information handling is twofold. First, the information-handling capacity of the small firm is a finite resource and a factor in the assessment of reasonable practicability basic to UK health and safety law. Second, styles and patterns of information flow are of consequence to the effectiveness of regulatory interventions.

This report and the work it describes were funded by the Health & Safety Executive. Its contents, including any opinions and/or conclusions expressed, are those of the author alone and do not reflect HSE policy.

HSE BOOKS

First published 2001

ISBN 0 7176 1990 7

Summary

The purpose of this study is to consider the possible advantages of an approach to regulation that concentrates upon the information handling patterns of small firms. The general orientation accords with the Robens philosophy of self-regulation and attempts to explore the HSE's role as a designer of self-regulation in small firms. This research does not review HSE's current practices and confines itself to the value that a different set of concepts might offer to HSE's work.

This paper presents a simplified view of a cybernetic approach to regulation. Cybernetics is a branch of systems theory that is concerned with how control is achieved in systems and is designed to address dynamic complexity. This considers regulation as essentially a problem of directing and processing information. Agencies such as HSE recognise the financial pressures on small firms and take account of them in judgements of regulatory impact. However, less visible are the constraints of information processing and these, though less visible than financial resources, are argued to be at least as significant. The importance of information handling is twofold. First, the information-handling capacity of the small firm can be considered a finite resource, a factor in the assessment of reasonable practicability basic to UK health and safety law. Second, the styles and patterns of information flow within the business and between the business and its environment are argued to be consequential in the effectiveness of regulatory interventions.

Three case studies were undertaken in the course of this work. Their purpose was to provide practical illustrations of the principles under discussion. Further insight into the bearing of cybernetics ideas upon the HSE regulatory task was gained by the partnership of the researcher with a FOD inspector and the integration of the visits into statutory inspections.

Many of the practical suggestions made in this report are already present in HSE regulatory practices, what is new is the conceptual framework in which they are ordered and discussed. It is this set of concepts that form the significant deliverable of this study.

Development of an Information-based Approach to Self-regulation of Health & Safety in Small Firms

John Kingston-Howlett
John Kingston Associates
Worcestershire, UK

Contents

1 INTRODUCTION

The Regulation of Complex Situations

The task of regulating health and safety is complex, complex in two senses. First, in the sense of the mass of factual *detail* that is relevant (e.g. technical, academic and administrative categories of knowledge). Second, complexity as the *dynamic* interrelationship of elements within a system. Even systems composed of just a few elements are capable of considerable complexity in this second sense. In H&S, the systems concerned—whether scaled at a single employer or the whole of society—are hugely complex and the task of achieving regulation in them inevitably must deal with the issue of complexity. This report discusses the issues of health and safety regulation using the perspective of systems theory, a literature and set of approaches designed to address dynamic complexity. Particular emphasis is placed on cybernetics, a branch of systems theory that is concerned with how control is achieved in systems, whether they are natural or artificial. In the last decade, the use of the term cybernetics has been greatest in that sub-class of artificial systems: computers. This paper uses cybernetics in its general sense, that is, cybernetics as the science of steersmanship (from which, via Greek, the word was derived). By analogy, cybernetics is concerned with the steering of complex systems towards their goals in the face of difficulties.

A dominant theme in systems theory is the construction of models and this itself is a means of dealing with complexity– simplifying matters to manageable proportions. A common concern to people is the extent to which a model in use represents all of the important attributes of the situation of interest; in other words, oversimplification may remove important aspects of a situation from consideration. What makes a *true* model depends upon the perspective of the person who is making the comparison between the model and the real system. Different people may see the same system differently. In fact, it is misleading to assume that there is some ultimate viewpoint that could reconcile contrasting views. This is another reason why models are valuable: they are simpler and more comprehensible than the real world they depict and allow people to share alternative views.

However, obtaining these benefits partly depends on being able to construct models that relate to the real world in specific ways. This is clearly a requirement between a road map and the neighbourhood because the whole point of a road map is to plan a journey in an unfamiliar place. In the same way, models of systems can be greatly useful to planning one's interaction with them.

This paper serves as an introduction to cybernetics as applied to the task of regulating the health and safety performance of industry. In doing so, one means of modelling regulatory situations is presented.

As a whole, this report aims to help the exchange of information between different viewpoints and, in particular, provide a means to:

- bring into focus disparate attributes of HSE's regulatory task;
- connect different strands of FOD research and compliance initiatives;
- facilitate discussions between contrasting viewpoints in HSE and others involved in delivering safe and healthy work;
- share information about the informational aspects of particular dutyholders or categories of dutyholder:

The ideas presented here are advocated also as a basis for planning interventions:

- to allow different levels of intervention (e.g. strategic and tactical levels) to be discussed in the same terms and interrelated.

- a means for designing interventions tailored to the informational characteristics of dutyholders;

This paper uses the word *regulation* as a verb—as in self-regulation. For the purposes of this work *regulation* means: any systematic behaviour of one part of a system that tends to restrict the fluctuations in behaviour of another part of that system.

It is important to be clear that a "regulator" is not an identity but a role, or function of an element in a system. As such, the regulatory role of an element may not be its only function in the system. Even more important is to keep in view that the regulatory character of the behaviour of one element very much arises from the relationship between two parts rather than the identity of one of them.

This accent on the affirmative active sense of regulation is much in keeping with the Robens' view of self-regulation as actively involving all people levels of industry. Ultimately, the resources used to achieve HSE's mission are primarily those of the dutyholder and, with this in mind, the Robens credo is more easily accepted for large-scale enterprises than it is for small businesses.

That small firms have less resources and present special difficulties is easy to accept. Although money is the most obvious resource that is both relevant and limited in small firms, it is nonetheless reasonably flexible. Perhaps an even more pressing limitation is the information resources of the small firm: not just availability of information but, more importantly, the information processing resources that are limited by the number of people and time. This paper emphasises the importance of accounting for information resources when designing interventions with small firms.

Cybernetics sees regulation as an essentially informational process, one in which regulation is achieved by operating a network or a hierarchy of communication. Within this perspective, HSE achieves its mission through both the *regulation of dutyholders* and the *design of regulators* at every level of the regulatory hierarchy including the dutyholder.

1.1 Robens Self-regulation in the context of systems theory

The view of regulation described here arises from the *cybernetics* and *General Systems* literature. Heylighen (1997) characterises this as

> "Cybernetics and General Systems Theory ("Cybernetics" or "Systems Research") are two domains which have undergone so much cross-fertilisation that they are in practice difficult to separate. They aim to study all complex systems, independent of type or components. They focus on concepts such as boundaries, input-output, state spaces, feedback, control, information, hierarchy and networks".

Fundamentally, the systems perspective sees regulation as an informational process; one in which communication and control are intimately related. When the informational character of regulation is emphasised, various insights emerge. For example, within this perspective a dutyholder's information resources are as a relevant as his financial situation when weighing-up the balance of costs and risks within SFARP. Furthermore, the ways in which a firm acquires information and uses it are relevant considerations in the design of regulatory interventions. The text explains these ideas.

The cybernetics perspective is consistent with the world as seen by Robens:

> "The most fundamental conclusion to which our investigations have led us is this. There are severe practical limits on the extent to which progressively better standards of safety and health at work can he brought about through negative regulation by external agencies. We need a more effectively self-regulating system. This calls for the acceptance and exercise of appropriate responsibilities at all levels within industry and commerce. It calls for better systems of safety organisation, for more management initiatives, and for more involvement of workpeople themselves. The objectives of future policy must therefore include not only

> increasing the effectiveness of the state's contribution to safety and health at work but also, and more importantly, creating the conditions for more effective self-regulation." (HMSO, 1972; paragraph 41)

Robens dealt with the macro level arrangements that he saw as the prerequisite to self-regulation. Examples include the unification of the inspectorates, the embodiment of a principle of assent in the HSC, the simplification of health and safety law. What is left mostly unsaid by Robens is how to connect the macro level to the micro-level (i.e. the operating level of industry) to achieve the goals of the new policy.

Implicit in the Robens' report but explicit in cybernetics is recognition of the informational resources required in the regulation of a complex system by a regulator of (relatively) limited capacity. There are many principles applied to achieve this, but among them is the central notion of utilising resources outside the regulator to achieve effective regulation. Correspondingly, the issue before and since Robens is how to harness the resources of duty-holders to the H&S goals indicated by the Law, the Government and its agencies. FOD already practises in this philosophy, as demonstrated by innovative compliance strategies such as TORCH.

To put this another way, the status of self-regulation as a credible philosophy is not in question; no dutyholder undertaking would last a 100 days without internal mechanisms of control. The question then is how to secure the ambitions of HSE policy through the mechanisms of control that exist in society.

Part of the cybernetics perspective is that the problem of control in goal-oriented systems (like organisations) is amenable to a *hierarchy of regulation* (Aulin, 1982 & 1989). The word *hierarchy* carries connotations of command and control and of prolific "ordering-around", this interpretation is to be avoided. For our purposes, hierarchy refers to a logical precedence of choices; choices that ultimately determine the characteristics of the operational system at the "bottom", as it were, of the hierarchy. These hierarchies are, therefore, primarily informative and their structure is revealed by the passage of information between the different parties involved. Many such regulatory hierarchies impinge upon a given real-world entity such as a corporate duty-holder, an employee, or a representative body. All of these create an intricate web of information that influence choices made at the operational level of the economy (meaning, the production and exchange of goods and services).

Robens emphasised the role of people as responsible decision-takers, and not just those individuals who wear the mantle of management but anyone in employment with discretion about how and when to act. The accent on regulation through people is a hallmark of the approach taken here.

1.2 Information, data and Knowledge

This perspective on regulation relies on a particular definition of information that emphasises the *active processes* of communication. These definitions of information are compatible with communication theory. Communication theory, developed in the 1940's, is a branch of mathematics dealing with the description of systems, especially the dynamic relations of systems. The generalisation of communication theory to regulation was achieved by Ross Ashby in the 1950's in a programme of research and writing that remains current in present day research (Heylighen, 1997). What is important is that the concepts presented here are supported have been developed with rigour.

Data will be taken to mean ***symbols which have not yet been interpreted*** (van der Spek, 1997). In the abstract, data are *statements of fact* (Beer, 1985).

Information refers to *data that have been assigned a meaning* (van der Spek, 1997). More straightforwardly, information may be defined as the portion of data which makes a difference; as Beer (1985; p. 283) has it, ***Information [is] that which changes us – data become information when the fact in them is susceptible to action***. Insofar as information can be considered in the abstract, it is *always linked to a specific situation and has only limited validity* (van der Spek, 1997).

Knowledge is what ***enables people to assign meaning to data and thereby generate information***. *It is the whole set of insights, experiences and procedures which are considered correct and true, and which therefore guide people's thoughts, behaviour and communication. Knowledge is always applicable in several situations and over a relatively long period of time* (van der Spek, 1997).

Information when seen as a facet of an *active* process is a useful discipline for communicators. It not merely requires that a message about a topic be *sent*, but that the communication is *designed* to achieve a change in the recipient. If the communicator acts in the role of regulator, the change in the recipient is a matter that will normally be verifiable by feedback. The upshot of this is that the regulator, as a recipient of data, is open to change (i.e. informed through feedback).

In considering the role of information and communication technology in achieving social improvements within the EU, the authors of a report to the European Commission make similar distinctions:

> *First and foremost it is essential to make a clear distinction between data, information and knowledge. From our perspective, the generation of unstructured data does not automatically lead to the creation of information, nor can all information be equated with knowledge. All information can be classified, analysed and reflected upon and otherwise processed to generate knowledge. Both data and information, in this sense, are comparable to the raw materials industry processes into commodities.*

The authors of the report go on to note that:

> "At the same time… it is essential to point out the difference between the production and use of tangible raw materials and intangible information (data): the latter can be reproduced at little cost to the producer."
>
> (European Commission, 1997)

1.3 Responsibility

The idea that regulation can be treated as an informational process lends itself to the socio-technical systems that HSE works with. To illustrate this at the level of the individual, these ideas are examined in terms of personal responsibility. From this perspective, the effect of regulatory hierarchies is to influence choices made at the operating level (of the workplace).

Central to UK H&S law, is the concept of individual responsibility. This is emphasised throughout the Robens Report. In current HSE publications (such as Successful Health and Safety Management, HSE 1997) it is recognised as an important issue in compliance, especially compliance when considered in terms of self-regulation. However, the analysis of responsibility from the cybernetic viewpoint means making explicit the role of information within a system with the "responsible" person as a focus.

Shaw & Barry (1989), consider Responsibility to have three distinct but inter-related meanings; shown below in Table 1.

Sense One **Responsible as in *account-able for past actions***	***Sense Two*** **Responsible as in *having a duty* (responsible for some-body or something of value)**	***Sense Three*** **Responsible as in *rational (able to make informed decisions and to act)***
To behave within given con-straints or beyond some mini-mum standard (as informed by law, an organisation, or soci-ety).	To have a role derived from the specific requirements communi-cated to the actor by higher au-thority	In this sense, responsibility re-quires requisite: (1) knowledge of salient facts (2) cognitive ability and (3) the ability to act.

Table 1 Three senses of *Responsibility*

Within this scheme, the various senses of responsibility are interrelated; Sense two [duty] requires Sense one [accountability] to provide some basis for judgement (i.e. duty without the qualification of accountability might be referred to as "notional responsibility"). However, the real importance of this is that, whilst each sense of "responsibility" is distinct, the scheme is dominated by Sense three [rational]. If not responsible in Sense three, one cannot be responsible in senses one and two and, as can be seen, this presupposes a requisite state of information in the individual.

Phrasing this in terms of the HSE's situation, *the law underpins sense two* [duty], it prescribes the duties of people at work for the health and safety of people. Equally important is the fine-tuning of duties by the employing organisation to accommodate the particular features (work activities and associated hazards) of the work setting. All of this presupposes the active processes of communication at each level with the emphasis on *downward* communication through a hierarchy. It is worth noting, as Beer (1985) does, that the *emphasis* may be downward but, in practice, the communication will resemble more a negotiation to reach a bargain than a dictat summarily delivered. In HSE terms, this is more compliance than enforcement.

Sense one [accountability] is the reciprocal of sense two [duty]. Sense two, as the assignment or acceptance of a duty, requires that its discharge is a matter of consequence– the dutyholder can be held accountable. As such, it implies communication at each level with the emphasis on *upward* communication through a hierarchy. **Sense three** [rationality] will be specific to the particular work system that the dutyholder is expected to manage, to do this requires:

- data about the current state of the work system;
- a mental model of the work system;
- the ability to manipulate the mental model;
- a set of values to inform decision-making;
- the ability to implement the decision - to act;

Three things are clear in this account of responsibility:

- the same considerations apply at any level in the hierarchy;
- at any given level of the hierarchy, the involvement of the next level "up" is implied;
- channels of communication bind the whole hierarchy together.

The remainder of Section 1 explores the means of generalising the ideas illustrated here and introduces the further cybernetics principles necessary to do so.

1.4 Models

The result of a simplification can be called a model: the modeller has, as it were, scaled-down a real-world object or system. There might be various reasons for doing this: because it is not feasible to experiment upon the real world or because the real world is too complex, too unwieldy to understand directly. In physical models the modeller reduces the scale but preserves proportions. Often the modeller explicitly states the scale of the model to the original. Various other attributes of the original are also preserved but which they are depends upon the modellers purposes. For example, a wind-tunnel model preserves the surface contours of the original because these are consequential to predicting the aerodynamic behaviour of the original. On the other hand, attributes of the original such as its colour, internal features and size may well be lost because they are inconsequential to its aerodynamic performance. On this basis, a good model is one that fulfils the purposes of the designer of the model. Similarly, the validity of a model is the extent to which it preserves features of the real-world object or system that are consequential to these purposes.

1.4.1 Requisite Variety

A limitation encountered in analysis that emphasises the informational characteristics of regulation is the loss of some of the richness and colour associated with social modes of explanation (like motivation, politics, and ethics). This is one of the costs of simplification and it needs to be weighed against the benefits of concise identification of situations that require and invite different interpretations on how to intervene.

Whenever a complex situation is modelled, the modeller needs to find a balance between functionality and usability in the model: too complex and the model is unwieldy; too simple and consequential details are not preserved and the model lacks functionality. Another way of expressing this is provided by Ross Ashby who suggests that a measure of the complexity of a system is *the number of distinguishable states that can be expressed in that system*. This quantity he terms ***variety***. For example, throwing a dice can have six distinguishable outcomes; its variety is six. Applying the concept, a model that contains all the relevant features of the reality it depicts has *requisite variety:* all consequential distinguishable states of the real-world system map onto the model.

Returning now to the issue of regulation, the model (whether internally or externally represented) of the system regulated is subject to the same issues of variety. The regulatory model, as any model, includes and excludes features of the real world as dictated by the modeller's purposes. These purposes determine which features of the real system are relevant to include in the model. The extent to which he is successful in incorporating all consequential features into the model is a highly relevant factor in determining the overall success of the regulation of the real-world system. Consequential states of the real world that are not distinguishable in the corresponding regulatory model simply do not exist for the regulator. This, and its mathematical proof, is the Ashby-Conant maxim (Conant & Ashby, 1970):

> *Every good regulator of a system must be a model of that system*

The significance of the maxim is perhaps even greater. Within the systemic view of regulation, the regulatory model is *equivalent* to the real world system: all regulation is determined by actions based on how the real world is reflected in the model. Therefore, when an inspector is considering what to do and how to do it within a workplace, the success of the action is largely determined by the adequacy of the regulatory model that s/he develops of the dutyholder work-system. The same argument applies at higher and lower levels. HSE bases its strategic decisions on models of what it believes the real world to be. This is also true of are actions and measurements of effectiveness (like inspection numbers, notices, prosecutions etc). At the lower level, an operator has a regulatory function that is dependent upon the model they maintain. If this model lacks variety with respect to health hazards, precautionary action is unlikely (unless insisted upon at a higher level, in which case the regulatory role and model resides at this higher level in this respect).

1.5 Conventions for describing regulation – D, T, E & R

Ashby provides a simple scheme for exploring regulation that allows the various aspects of the regulatory relationship to be explored in terms of four functional elements: ***Essential variables, Disturbances, Transformation*** (e.g. the operational aspect of a work system) and, ***Regulator***.

- **(E) Essential variables** are those aspects of system performance that are important and, in an organism, that are essential to viability. For each essential variable there will be a critical range of values that are acceptable and outside of which the system must not deviate.
- **(h).** Considered jointly, the acceptable ranges of the system's Essential variables form a set denoted by the letter "η" (eta). What is wanted is to keep the system performance within η and this is the task for regulation.
- **(D) Disturbances** are any influences that tend to push the values of essential variables outside of the eta set of acceptable values.
- **(T)** Stands for **"Table of Transformations"**, an element that takes inputs and transforms them to outputs in a systematic way.
- **(R) The regulator** modifies the workings of T in the light of Disturbances to eta. These disturbances may be communicated to R by the movement of essential variables beyond the eta set of acceptable values **("error controlled").** Alternatively, the occurrence of D may trigger a regulatory response in R, before D affects the Essential variables **("cause controlled").**

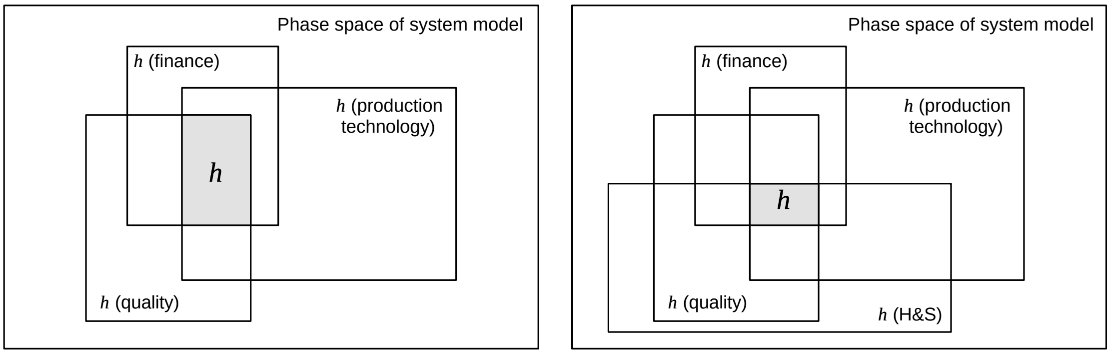
1.6 Relationships between E, h, D, T & R

1.6.1 Essential Variables and the set of acceptable values (eta)

How essential variables are established in a specific instance is often a complex matter. However, in general, what counts as important derives from the purposes of the work system.

The effect of selecting a range of acceptable values of an essential variable is to constrain what the work system does and how it does it. Hence, say a unit production cost is accepted in a manufacturing process; this will have the effect of constraining operational freedom. Similarly, when a quality standard is applied, this too reduces operational choice. What remains, when all the essential variables have been applied, is the set of desirable states of the system that operations are at liberty to choose from to achieve the operational work of the system. Ashby denotes this as η (eta). If we consider the unconstrained work system in terms of variety, that is, the number of distinguishable states, the application of E is to reduce this variety. Of the huge variety things that the work system could be used for, only a subset will be pursued. Further, of the large variety of ways in which that subset could be achieved by way of method, people, materials and technology; again, only a small proportion will be accepted.

Figure 1 illustrates the idea of each Essential variable removing degrees of freedom from a system. As a particular range of values is asserted, the number of acceptable configurations of the system becomes smaller. As each essential variable is factored-in, the operational freedom of the system is further restricted. Considered on its own, the impact of any one essential variable upon operational freedom may not seem too great. However, the impact of all the essential variables is likely to be considerable.



Operational freedom without inclusion of H&S essential variables

Operational freedom with inclusion of H&S essential variables

Figure 1 Constraint of operational freedom by application of performance criteria

One point here is that the reduction of operational freedom may seem oppressive. Whether it actually is perceived that way depends upon various factors but, importantly, issues of motivation. The constraint of operational freedom is often explored as a motivational topic in health and safety. When looking at the level of the HSC, the consensual process is one of reaching a shared vision of purpose between Industry, Unions and Government. Within the small business, inspectors may follow a similar logic in defining boundaries that balance operational freedom with the essential variables of health and safety.

The DTER formulation is very general, necessarily so if we are to connect the highest with the lowest levels of regulation. As the firm is part of the society to which it belongs and the industry of which it is a part, some essential variables are subject to wider influence and, hence, regulation at a higher (logically higher) level. In the case of assets that belong to society such as the environment and individuals, the state has a duty to ensure their wellbeing and preservation. However, government has to balance the benefits of industry to society with the costs of achieving these gains. In other words, workpeople are the industrial means by which society achieves its ends. Whilst not considered further in this report, this indicates the connection with the wider issue of how the State arrives at the η set: the "what is wanted" performance to be achieved by industry. From HSE's perspective, most of the state's regulatory resource is to be found in industrial society as underlined in the Robens' notion of self-regulation.

1.6.2 Regulation and h

In work systems, essential variables are maintained with desired limits by processes that regulate the exchange between elements of the work system and between the work system and the environment. Exchanges that have a negative impact upon the essential variables of the system are called Disturbances. The purpose of the Regulator is then, to ensure that disturbances impinging upon the work system do not force upon the values essential variables out of the ηset

With this in mind, regulation can be seen as the selective blocking of information (incidentally, an illustration of the definition of information as "that which changes the recipient"). This is most easily be imagined as a physical blocking, such as a pair of suitably specified ear-defenders or an insulated protective suit for working in heat. Alternatively, it may be a blocking effect where the information impinging upon the system is opposed by information generated by the regulator. In the latter case the regulator may cause changes in (i.e. inform) the work system to counteract the changes originating from the disturbance. What changes little (or not at all if regulation is perfect) is the values of the essential variables; these stay with the η set.

An example might be the action of a defender in basketball to the basket shooting attempts of an opposing member of the team. Each duck and dive of the would-be scorer is matched by the defender blocking the transmission of the ball to the basket. One of the case studies undertaken to illustrate this paper, was a bakery. There, the baker would act to regulate the quality of his products by countering disturbances such as variations in the gluten content of flour, the saltiness of meat for pies.

Similarly, the ear-defenders mentioned above which achieve their effectiveness by a mixture of reflection and absorption of acoustic energy have their equivalent in the use of anti-phase acoustic techniques. Here, the acoustic signal arriving from the environment is simultaneously received and a new signal (the phase of which has been adjusted to cancel out the original signal) sent to the system.

Implicit in all of the examples provided (particularly evident in the last case) is the need for the regulator to match <u>message for message</u> the environmental disturbance that would otherwise adversely impinge upon the essential variables of the system. Hence current anti-phase technology is extremely effective when the noise is highly predictable but has problems when noise is unpredictable in any of its parameters (Fitzgerald, 1996).

Ashby provides a simple scheme for describing the relationship between the disturbance, essential variables, work system and regulator. Within this scheme, what we have been calling the *work system* is denoted *T*. As mentioned earlier, T stands for "table of transformations" and is a general way of referring to the element in a system that interacts with its environment to process inputs into useful outputs that achieve the purpose of the system. T, in essence, does what the system exists to do.

How we define the boundaries of the system is a matter of choice and it is helpful to remember that T could be a lathe and the skills of its operator producing finished components for a customer. T might as easily be the whole set of operational activities with the firm. It is important to be consistent once a particular focus of a system is taken. Also, one should always bear in mind that the different elements D, T, E and R do not necessarily share boundaries with real-world entities or objects in the system. For example, a small businessman may be simultaneously on-the-tools (T), ensuring the right quality and quantity of a product (E), deciding as he goes, when to change the cutting bit, when to take a break, how to accommodate any unusual qualities in the materials (R), but bought cheaper materials with more variable qualities, running the machine faster creating greater wear of the cutting bit and errors in his machining (D).

1.6.2.1 Regulation in terms of D, T, E & R

Using the terms established, Ashby (1956) provides a simple formulation of the regulatory context:

"In practice the question of regulation usually arises in this way: The essential variables E are given, and also given is the set of states h in which they must be maintained if the organism is to survive (or the industrial plant to run satisfactorily) These two must be given before all else. Before any regulation can be undertaken or even discussed, we must know what is important and what is wanted."

To put this another way, when you know what your essential variables are and the range that they must be held within, the transformation taking place and the negative impacts on this transformation, it is then possible to establish regulation on the transformation so that the essential variables are maintained within the necessary business survival limits.

Ashby (ibid., page 220) provides a simple description of the regulatory task as:

> *"Given E, h, T, and D, to form the mechanism R so that R and T, coupled, act to keep E within h."*

There are various different ways in which these elements can be joined together (i.e. in the sense of brought into communication), the chief variation being the route by which R is informed about D. The simplest and most direct is by D itself: this is known as cause-controlled or feedforward regulation (further explanation given at section 1.6.2.2, page10). The intriguing aspect of regulation that the cause-controlled type emphasises is that R achieves its regulation of T by reacting to D variety in a manner that absolutely mirrors the reaction of T to D. This emphasises the Ashby-Conant maxim: *Every good regulator of a system must be a model of that system; R's behaviour is equivalent to T's given D.* The same argument underlies the regulatory principle of requisite variety, as Ashby states it:

> "only variety in R [the regulator] can force down the variety [in the essential variables of the system] due to D [the disturbance]; **only variety can destroy variety".**

Ashby, 1956

1.6.2.2 Simple regulatory configurations of D,T, E & R

Whilst the basic principle enunciated above is common to all regulation, there are two main forms, **cause-controlled** regulation and **error-*controlled***. Figure 2 illustrates cause-controlled regulation as a directed graph; this shows the flow of data between the various elements.

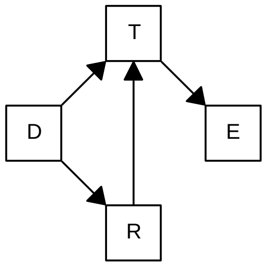


Figure 2 Cause-controlled regulation
(NB. The arrows indicate data flow)

The cause-controlled regulator is, in principal, the superior design because it can act in advance of any adverse effect on the essential variables. Figure 3 shows the information route more clearly. The term *feedforward*, is sometimes applied to this scheme.

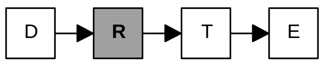


Figure 3 Intervention by R in the Cause-controlled paradigm

The cause-controlled regulation scheme affords *R* the possibility of perfectly blocking the flow of information *D* to *E* (subject to *R's* variety relative to *D* as mentioned previously). Communication theory makes clear that the regulator can only block disturbance from being passed on through T to E up to a limit: this limit being the variety of R itself. Extra variety of disturbance will "get past" R and may transmitted onto the Essential variables. In terms of risk management, this margin of control is residual risk. In this respect, the relationship between communication theory (and particularly, Shannon's Theorem 10–the "the noisy channel theorem") and Ashby's requisite variety is exact (Atlan, 1983, provides a proof of the formal correspondence between this theorem and requisite variety).

Cause-controlled regulation looks attractive, at least initially, when viewed from the perspective of health and safety management. R acts before a hazard is realised at E by making changes to T (the operational system): definitive proaction. However, this poses some practical problems: how to construct R in the first instance (i.e. complete with this response), and how to keep it effective if D increases in variety or η is changed.

Error controlled regulation is required when the regulator cannot react directly to a disturbance, for example because it does not have sensory connection to D and hence no channel capacity with respect to D. Figure 4 illustrates this as directed graphs.

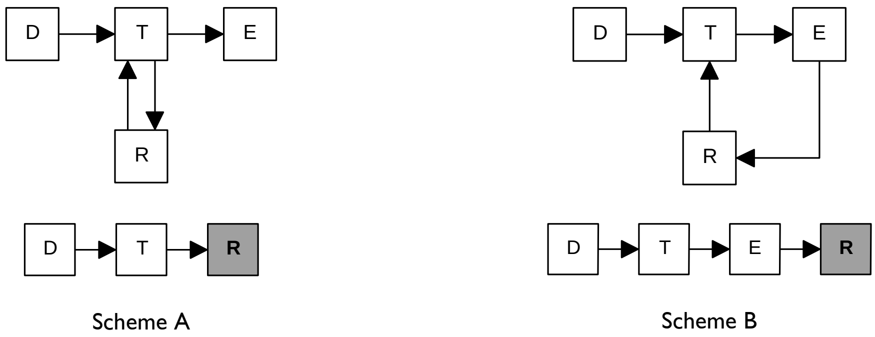


Figure 4 Two generic forms of error-controlled feedback

Under scheme A, as the lower portion shows, the regulator obtains its information by way of *T*. This means that the disturbance is already acting on *T* before the regulator can begin to act. Additionally, T may not perfectly preserve the information content. For these reasons, the error-controlled regulation can not be made perfect in the sense that cause-controlled regulation can. Were the regulator intelligent (e.g. human) this scheme suggests that *R* can only make sense of what is going on (the action and nature of *D*) through, as it were, the eyes of *T*.

Scheme B takes matters yet further. Because *R* is receiving its information via *E*, successful regulation (e.g. through trial and error) means that the channel of information from *D* is blocked to some extent. As Ashby puts it: *"the more successful R is in keeping E constant, the more does R block the channel by which it is receiving its necessary information. Clearly, any success by R can at best be partial"* (1956, page 224). However, so long as it is true that the effect of the disturbance is not so dramatic as to force the *E* value(s) outside of their viable range before *R* can act, the scheme can prove highly effective. Small errors are compensated by regulatory action thus providing adaptation to the *T-E-R* complex overall with respect to the environmental variations *D*.

Whilst this seems a much less efficient means of achieving regulation, it is simple. A detailed algorithm that predicts every state of *D* and plans every response by R need not be prepared. The system's behaviour can be characterised as never reaching equilibrium until the values of the essential variables are within the set h. From the perspective of R: until the essential variables (E) are within the range of values that are acceptable (η), R will keep hunting (in the engineering and deer-stalking senses) for an set of inputs to T that will remedy the situation and then "stick". If R has a memory, whenever the D situation occurs, the new regulatory response will be available… R has adapted. Ashby refers to this as regulation by vetoer or *"hunt and stick"*.

Many systems are likely to have elements of both cause and error-controlled regulation, each method deployed to different states of D. The combination of both forms, i.e. both cause- and error-controlled modes of regulation, is referred to as the *full information* paradigm. Within Health and safety management, the preference is for proactive, cause-controlled regulation. However, this may require greater variety generation than exists in the firm. Even in the largest of organisations, cause-controlled regulation is never perfect and, indeed, the fact that safety case risk assessments do not predict the form of later accidents attests to this limitation. Therefore, although optimisation of cause-controlled regulation is clearly desirable and, for some hazard forms, the only means of achieving requisite variety as defined by SFARP (So Far As is Reasonably Practicable), there are various limits. Error-controlled regulation is highly effective, relevant to adaptation and *economical* to establish.

1.7 Preserving Requisite Variety and the design of regulators by regulators

Ashby stated requisite variety as a *law*, meaning the relationship between its constituent arguments is *regular*. The law of requisite variety will always assert itself. For example, if a regulator has just enough variety to maintain essential variables within η, an increase in the variety of disturbances will outstrip these defences. Similarly, if the variety of R and D are held constant but a narrower range of values accepted for the essential variables, R will need greater variety defend η. The point here is that with regulatory effectiveness as a goal for work system design, one must recognise that requisite variety is not a property of the regulator alone but a function of all the elements in the regulatory picture. Hence, while it is intuitive to concentrate attention on building an effective regulator– to place R centre-stage–we need to stay alert to the <u>systemic</u> nature of regulation.

Regulation can exploit four options, whether the system is large or small:

1. Allow a wider range of values in η;
2. Reduce the variety of the Disturbance;
3. Exploit constraints in the variety of Disturbance;
4. Increase the variety of the Regulator.

1.7.1 Allow a wider range of values in h

This, in essence, is to lower the standards as far as viability allows. The phase space diagrams of Figure 1 (page 8) illustrate the impact of several classes of essential variable: η becomes very small – and restricts operational freedom. By analogy, the regulatory target becomes smaller and harder to hit.

As Rasmussen (1996) and Senge (1992) note, the expansion of the η set boundary to allow greater operational freedom, is a generalised problem (i.e. across different categories of adversity). Further, both authors note that this phenomenon is manifest as gradual change that is unopposed, "creeping quality" rather than an explicit decision to accept a lower standard. Rasmussen refers to *entropic drift*, and Senge– *eroding goals*. Rasmussen provides a complementary perspective of this issue, he argues for the:

"… Existence of a natural migration of activities toward the boundary of acceptable performance and we have to find a way to represent the mechanism underlying this migration"

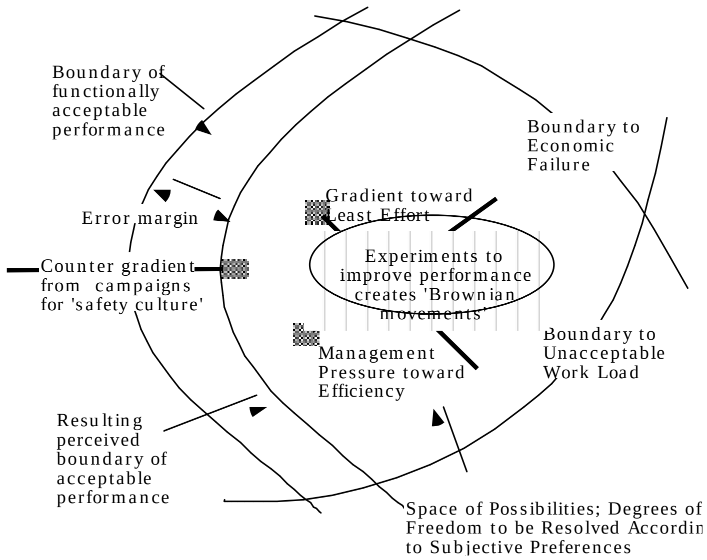


Figure 5. From Rasmussen (1996) "Under the presence of strong gradients behaviour will very likely migrate toward the boundary of acceptable performance"

" Human behaviour in any work system is shaped by objectives and constraints which must be respected by the actors for work performance to be successful. Aiming at such productive targets, however, many degrees of freedom are left open which will have to be closed by the individual actor by an adaptive search guided by process criteria such as work load, cost effectiveness, risk of failure, joy of exploration, etc. The workspace within which the human actors can navigate freely during this search is bounded by administrative, functional, and safety related constraints. The normal changes found in local work conditions lead to frequent modifications of strategies and activity will show great variability. Such local, situation-induced variations within the work space call to mind the "Brownian movements" of the molecules of a gas. During the adaptive search, the actors have ample opportunity to identify "an effort gradient" and management will normally supply an effective "cost gradient." The result will very likely be a systematic migration toward the boundary of functionally acceptable performance and, if crossing the boundary is irreversible, an error or an accident may occur". Rasmussen (1996)

Hence, the option to achieve regulation by "allowing a wider range of values in" is one that naturally occurs. Even though this is likely to be indefensible as a managerial option for health and safety, that does not mean that it is not used. Firms may unwittingly take advantage of this effect to achieve operational ends without incurring the costs associated with improving other parts of the regulatory equation (i.e. options 2-4 above).

1.7.2 Reduce the variety of the Disturbance

This is often the first option within H&S regulation and, insofar, as we are addressing man-made systems, it is quite apparent that to "remove or reduce the energy" is the top of the safety control hierarchy. However, there are three limits to this approach.

1. First, operational effectiveness will require a certain quantity of variety and will seek to preserve it. Some proportion of the total variety of disturbances impinging upon a subsystem will originate from it and other subsystems within the system as a whole.
2. The second limit is the power of an entity to reduce the variety of the disturbance. For example, the regulatory approach before the HSW act was to limit or proscribe certain processes and substances (the use of Asbestos in Europe and the US is a prime example). The state as a regulator has practical limitations on its own variety, expressed both in acquiring adequate information about industry and handling information through so constrained a channel as the parliamentary process. This is in addition to the political difficulties created by challenging the operational variety that industry will seek to preserve.
3. The third limit concerns the portion of *D* that truly arises from the environment beyond the control of the system designers.

1.7.3 Exploit constraints in the variety of Disturbance

The point here is that although the variety in the components that comprise *D* may add to a huge total variety, *D* may nonetheless show some form of constraint. In other words, the actual variety of *D* is less than estimated. The argument here is that if one cannot increase the variety of *R* (or *D* artificially decreased) we may have little option other than to regulate against the repetitive, or constrained, variety of D. Hence, a low-variety regulator such as the cough reflex will clear the most frequent portion of disturbance variety whilst leaving rare occurrences such as impaling fish-bones unregulated. A procedure for working on the roadway serves as an another illustration. The variety of D (e.g. different vehicles at various speeds and directions) will show constraint and can be controlled through a low-variety procedure– including, for example, the display of warning signs, cones, and the behavioural routines of the workers themselves. However, it will not be effective in all instances (e.g. drunk drivers, mechanical failures in vehicles, exceptional situations in the workers themselves– new, ill, responding to an accident, etc.).

The perspective here is of the *strategic* long run, regulating the repetitive disturbances and living with the effects of the rare non-repetitive disturbances. Ashby provides an engaging example:

> "It should be noticed that what is "Good" in the Grand Outcome [the long-run] does not necessarily follow from what is "good" (η) in the individual outcomes; it must be defined anew. Thus if I go in for a lottery and have three tickets, a win on one (and consequent loss on the other two) naturally counts as "Good" in the Grand Outcome; so here 1 good + 2 bad = Good. On the other hand, if I am tried three times

> for murder and am not found guilty for one, the individual results are still 1 good + 2 bad, but in this case the Grand Outcome must naturally count as Bad. In the case when the individual disturbances each threaten the organism with death, Good in the Grand Outcome must naturally correspond to 'good in every one of the individual outcomes'." (Ashby, 1956, page 249)

An ethical issue here is the acceptability of "bad" individual outcomes (people as citizens, as ends in themselves) as contrasted with a "good" grand-outcome for the organisation (people as work-force or local residents, as means to organisational ends). As HSE use both of these as bases in decision-making, one might a certain tension between those who typically make decisions based on the individual case (e.g. Inspectors in the field) and those who make decisions based on the grand-outcome (e.g. strategists and other policy-makers).

1.7.4 Increase the variety of the Regulator

This appears less open to ethical compromise than the options so far discussed. However, increasing the capacity of a given regulator is often difficult in practice. *Increasing the variety of a given regulator is likely to meet practical limitations long before it meets requisite variety.* We should always try to get the best from a regulator but not expect the regulator to go beyond limits that communication theory suggests to be absolute.

1.7.5 Regulator design and amplified regulation

Here, instead of attempting to increase the variety of a *particular* regulator we attempt to increase the variety of *overall* regulation deployed against disturbance. However, this is distinct from the options discussed above because this invokes the involvement of resources from outside of the DTER scheme.

Regulatory amplification is achieved in much the same way as power amplification. Like electrical amplification, amplification of regulatory variety is, in fact, a supplementation achieved in, minimally, *two stages*: the energy provided by the input and the supplementation of this energy from some abundant source to produce a more powerful output. Similarly, the law of requisite variety requires that a low-variety input be supplemented using an abundant source of variety to produce a higher (i.e. requisite) variety output.

Hence, a regulator with a variety of say 10 units, facing a disturbance of 1000 units must find an amplification of minimally 100x to reach requisite variety. This statement, whilst purely illustrative provides the necessary change of focus, namely, the view of regulatory amplification as involving design: the obtaining of a second (or indeed, third, fourth, etc.) stage of regulation. Page 5 introduced, in embryo, the idea of hierarchy "as a logical precedence of choices". The principal is that supplementation of choices is a means of increasing the variety of selection overall. Further, the fact that selection can often be achieved by stages carries with it the implication that the whole selection can be carried out by more than one selector. This means that the action of one selector can be supplemented by the action of others.

Whenever something complex is to built, it is common to divide the task into separate teams and to manage the programme centrally. This basic principal is very widely demonstrated, in industrial practice and in developmental biology to name just two diverse fields. If we view the construction of a complex system as an open-ended commitment– where the operational system is continually built-up against the buffeting of disturbance– *design* can be seen as very similar to *regulation*. The continual renewal, or self-perpetuation, of a system is termed *autopoesis*.

Within cybernetics, design and regulation are subject to precisely the same principles and treatments – different sides of the same coin. Part of learning to work with these ideas is developing flexibility with the equivalence between design and regulation. Much of regulation is achieved by designing a lower-order regulator or a set of them. By way of abstract example, the species is a higher (logical) order than the individual; the decisions about where to place individual bricks are of a lower order than the choice of what form the house should take.

By way of convenience, the different orders of regulator are here represented in the form R_n, where the relative order of the regulator is given by a number. In the recent examples: the species is R_1 to the R_0 level of the individual, the Architect's choices are at the R_{1+} order relative to the R_0 level decision-making of the Bricklayer.

Design need not be the mechanical massively feed-forward processes that come easily to mind in the context of major hazard process industries. It can also be the simple selection by the higher regulator of latent attributes in the lower regulator based on performance feedback from the operational system.

What is distinctive in the cybernetics approach to this matter is that design:

- does not necessarily entail the absolute specification and construction by R_1 of R_0;
- can be seen as selections of latent states of R_0 by R_1;
- can be brought about by R_1's selective reinforcement of certain of R_0's outcomes.

Furthermore, the great emphasis is on the design of communication at all levels of regulation, especially the feedback channels.

The issue of regulatory amplification in social systems is extensively studied by Aulin (e.g. Aulin, 1989). He uses the concept of self-steering to denote a regulator's variety. The idea is simply that, in social organisations, the individual capacity for self-steering tends to increase with experience and learning. As this capacity increases, the effectiveness of the regulation achieved by the individual increases. As the capacity of a lower order regulator increases, so decreases the number of stages of amplification required to achieve requisite variety overall. Using the analogy of power amplification, if you need 10kW of power output from an input of 1W and all you have are 10W amplifiers you need a series of four. However, if the lowest order amplifier can be increased to 1000W you need only two stages. The impact of this logic is present in the assumptions made about large firms: they are believed to embody a large stage of amplification to the input of H&S law and government guidance. Similarly, small firms have less capacity and need extra stages. However, in regulation the units are not electrical power but *informational power* – more easily termed as *the power of selection* (selection, as used above, in the sense of *choice*). Just as when the objective is to achieve requisite variety of regulation of a system, when we contemplate the idea of regulation in stages, the idea of requisite hierarchy is useful. Too many stages of regulation are wasteful and lower orders of regulator will be perceive the situation as oppressive. Too few, and the level of regulation may not be adequate to ensure the successful protection of η. The social concomitants for problems of requisite hierarchy are evident in everyday life. For example, demands for greater government regulation of a current disturbance perceived by the media to be acting upon the public, such as the price of goods, or food safety, or the integrity of pension schemes. Similarly, indignant noises as the yoke of State regulation bites into the hide of vigorously self-steering businesses.

Aulin argues that the self-steering capacity tends to increase at all levels of social systems. Straightforwardly, individuals are able to learn from experience and acquire knowledge via training and education that enables them to effect better strategies for improving their own attempts at regulation. Incidentally, we keep in sight of the fact that a person may embody more than one level of regulation but that they generally "wear different hats" when performing the different levels. The second part of this report uses examples drawn from small businesses. There, men and women occupy multiple levels of regulation: the strategist who produces the business plan that cuts down operational variety from the virtual infinity of a blank piece of paper to the level of variety that allows a bank to make an informed choice about the likelihood of a return on investment; the manager who divides a process into a inter-linked set of sub-processes each with its own regulator who provides the missing variety of choices to be made; and, very often, as an operator/craftsperson making the final choices, e.g. "this brick, placed in this way, with this amount of cement" or "this word, placed at this point in this sentence" etc.

Within the European Union, the need to increase the self-steering capacity at all levels of regulation is recognised by the Commission. Lifelong learning is a clear example of a fast changing technological environment: when the pace of change was slower, apprenticeships performed well in this role. The general rule is that any external source of supplementation will tend to increase the variety of a regulator. In the discussions of part two, the issue of supplementation is an important theme.

2 SELF-REGULATION OF HEALTH & SAFETY IN SMALL BUSINESSES– CASE STUDIES

Part I has introduced the basic ideas of this paper, what follows is an illustration of these ideas as applied to the informational aspects of regulation in small firms. The focus in this part remains on self-regulation of health and safety essential variables, but with increased acknowledgement of the general business context. The objective of this part is to help make the principles clear through demonstration in the familiar context of small firms. To this end, three firms were visited. It should be noted that the purpose of the visits was to provide real examples for this report and, as an additional gain, to obtain further insight into the application of the ideas under discussion to FOD's work with small firms.

2.1 Background

As part of this research, Ms Gill May, a field inspector stationed at Manchester area office, made three visits to different duty-holder premises with the author acting as an observer. The intention of visits was to produce illustrative materials for this report and to provide a means of gaining the insights of an active field inspector to this perspective on dutyholder compliance. These were made, in two cases, as normal statutory visits with no notice given to the duty holders. On one occasion, the visit was a follow-up call to a prohibition notice served on a firm after a serious accident. In all three cases, the statutory purposes of the visit took precedence over the goals of this research. The visits were took approximately two hours each, a little longer than the average inspection, but not so long as to be burdensome to the duty-holder is concerned.

The preparation for the visits involved a meeting between Ms May, Mr Graham Piggott and Mr Gordon Crick at HSE's Offices at Daniel House, Bootle. On this occasion, Ms May was introduced to the theoretical orientation of this work and her role was discussed. The goal of the meeting was to identify the means of serving the needs of the research by integrating with the statutory requirements of the visits.

2.2 Interview Format

The method of gaining information from the firms was semi-structured interviews aimed at forming a picture of the firms' sources and handling of information derived directly from the Ashbian scheme presented earlier. The actual technique used by Ms May was to integrate the questions indicated below into her own style of inspection. As indicated earlier, where a statutory issue became apparent, the research mode of the inspection was relegated while the problem was handled. The interviews were conducted at the workstations of the individuals concerned. The individuals chosen were the owner/director and between two and four of their craftspeople/operators. The choice of operators was made pragmatically: who was available and how much time was available bearing in mind the variable time requirements of the statutory aspects of the visits. The structure was as follows:

1) Gain, by conversation, an overview of the range of tasks performed by the individual
2) In relation to one task, in <u>general</u> rather H&S terms, ask:
    a) What can go wrong in the task
    b) Why (unless self-evident) is this a problem
    c) What might cause it to go wrong

d) What to do if/when it happens
e) How they know to do this
f) Of whom would they would seek advice if a new problem occurred that they weren't clear how to deal with (and to give an example)

3) Repeat (2) for another task, again seeking general rather than H&S issues
4) Select a <u>safety</u> hazard(s) of the task and, with this in mind; ask questions (a) to (e), prompting (a) or (b) to help the individual understand the direction of the question.
5) Select a health <u>hazard</u>(s) of the task and, with this in mind; ask questions (a) to (e), prompting (a) or (b) to help the individual understand the direction of the question.
6) Ask about the individual's training: when and what?
7) Ask about (external) sources of information, prompting as necessary:
   a) Trade magazines
   b) Trade and professional bodies
   c) TEC's etc.
   d) Government (local and central) agencies

2.3 Selection of premises

The first premises, an engineering firm (20 employees) was selected because, as it had been served with a prohibition notice following a serious accident, a longer visit was justified. Furthermore, we reasoned that this firm's acceptance of HSE's interest as legitimate and credible was more easily won than otherwise. This would allow the researchers the opportunity to practice the questioning technique in more forgiving circumstances than the "cold call" inspection situation.

In view of the special circumstances of the first visit, the other two case studies were selected using two criteria: the dutyholding businesses should be in different industries; should be representative of the normal visit in respect of the period since previous inspection and not have advance notice of this visit. It should be borne in mind that the purpose of these visits was far more illustrative than "scientific". Any follow-up work seeking to refine the technique would certainly require more scientifically acceptable sampling and control.

The remaining two case-studies involved a bakery (owner-managed, 20 staff: six in the bakery itself and the remainder servicing four shops and market stalls) and a Dental laboratory, also owner-managed employing 10 people.

On all three occasions, the format was the same. On arrival, we would meet with the owner/director of the firm. Ms May would make the introductions, indicating that the statutory purpose of the visit. The researcher was introduced as a colleague who was accompanying inspections to study how information provision to small firms might be improved. When introduced to individual members of staff, Ms May would in each case introduce herself and the researcher and state the purpose of the visit. The author was in possession of appropriate authorisation (Form LP21) to enter each premise with an inspector under S20(2)(c)(i) of the Health & Safety at Work Act.

2.4 Illustrations using the Case Studies

The analysis described in the second part of this report builds on the framework provided in Part 1 to categorise, evaluate and report on the ways in which the small firms in the case studies use information to regulate their business. Regulation in this analysis is far wider than health and safety; it includes all issues of control and responsibility in the business.

As already discussed, the elements of the Ashbian scheme to functions and may not map onto discrete real world entities and objects. In the systems with which this report is concerned, work systems, R and T may often be different functional aspects of the same individual. For example, the decision-making aspects of human-machine interface design emphasise the R component whereas more automatic skilled operations such as welding or machine use place the focus more on T.

Another issue is the application of the DTER scheme to systems at any scale. For example, in the case studies undertaken to illustrate this part, it might be useful to consider the whole management effort as R in relation to the collection of operations T, that accomplish what the firm is in business to do. At other times, we might want to focus at a lower level, such as the operational level with one worker providing regulation of one element, T of the firm's work. The disturbances that act at the level of a single operator will certainly be included in the wider set of disturbances with which the firm has to deal as a whole.

2.4.1 To what does R correspond?

In work systems, R will be mostly a function of a person or group of people in the workplace. Although our focus is on the human role in achieving regulation, we should be alert to the possibility of regulation through mechanical means (e.g. control systems).

It is also reasonable to expect businesses to achieve regulation in more than one stage. For example, the engineering firm had one level of regulation to devise work systems, select particular craftspeople and equipment, and organise how these are put to work day-to-day. We can call this level of regulation, R_1 and the whole work system, T_1. Similarly, the consequential aspects of the firm's performance, in Ashby's terms the essential variables (E_1) of the firm, correspond to a range of acceptable states or conditions in the firm that collectively may be termed η_1. Changes that have a negative impact upon η_1 can collectively be referred to as the set of disturbances or D_1. Using the notions of responsibility mentioned earlier; to have a duty means to be a regulator of η outcomes, a duty achieved through the manipulation of T in the light of D.

In many small businesses, R_1 is a role played by a single individual who must be able to ensure that the whole variety of disturbances do not push essential variables outside of the range of values accepted for them (η). Whilst simple in principle, the reality may be demanding; *disturbances* may be high variety (e.g. ranging from staff absences, faulty materials, health hazards to the strong pound and the price of diesel) as may be the *essential variables* to be governed (such as, bank overdraft, product quality, staff wellbeing and the amount of time spent with her family). The law of requisite variety states that the variety of regulation and the variety of situations to be regulated against are balanced. In general this will be met by a hierarchy of regulation, each lower order regulator amplifying its senior. In work organisations, much of the regulation to be achieved is obtained through delegation to those best placed to control the part of the operational system that is exposed to the subset of the disturbances. When we focus at the operational level or *zero order* of regulation, it is useful to use the suffix "0" as a token of the level of focus. At the operational level, the regulator(s) here is R_0 and the disturbances, D_0 (similarly, the local operational or work system is T_0 and the essential variables affected are E_0).

2.4.2 Perspectives on the hierarchy of regulators

The chain of responsibility in the firms studied is viewed as a hierarchy of regulators, where each level has responsibility for making choices that place limits on the activities carried out at the level below. Normally a regulator in this analysis is an individual in a position of responsibility. E.g.

R_2	$\rightarrow$	R_1	$\rightarrow$	R_0
Owner/manager		Works manager		Operative

In order for any of the regulators in the scheme above to make appropriate rational choices direct the level below, a state of information is required that has "requisite variety". For example R_1, the works manager needs to understand sufficient of the process the operative engaged in, the inputs to the process, the potential disturbances, the likelihood of the process output stepping outside the limits set, and the working competencies of R_0 as an individual. Various factors determine whether the works manager possesses this state of requisite variety. These include: his data processing capacity; his cognitive ability to absorb and apply the data (itself a function of the works managers mental model of the process) and, most apparently; the ability to spend time in terms of shared attention to the number of tasks for which he has responsibilities.

Overall, the task facing the boss of the small firm has been stated in the abstract: *"Given E, h, T, and D, to form the mechanism R so that R and T, coupled, act to keep E within h*." This formulation holds true whether R is achieved in one stage or by a number of inter-linked stages or within the one individual or by a number of people.

For a first order regulator R_1 (perhaps an owner manager) it may be simple as hiring a craftsperson skilled in delivering (via T_0) the desired quality of work (η_0) under the normal range of conditions (i.e. D_0). Effectively, R_1 is *amplifying her own variety* through R_0 and, reciprocally, R_0 is accepting the reduction in operational variety entailed by R_1's direction of what is to be done and to what quality (η_0). For example, the welder at the engineering workshop had served a lengthy apprenticeship which delivered him into the workplace not only able to weld faultless materials with perfect equipment under ideal conditions (i.e. a skilled T_0) but under a variety of conditions, equipment and materials (T_0 given D_0). In fact this welder grumbled wryly about the quality of welding rod he is expected to use, noting that if he had a free hand he would be using a better (needless to add, more expensive) rod. Nonetheless, he still manages to deliver satisfactory welded joints albeit that he has to "work harder at it".

One can easily extend the two-stage arrangement to three stages. By way of illustration, the dental laboratory, for certain essential variables, is governed by the department of health (certain dental preparations are deemed to fall within the requirements of medical devices legislation). This requires traceability of materials from manufacture to fitting by the dentist and has necessitated investment in ISO 9000 methods and supportive technology within dental laboratory. The outcome of this matches the general effect mentioned above: R_n (Dept of Health) is *amplified* through R_1 and, reciprocally, R_1 is accepting the reduction in operational variety entailed by R_n's direction. Interestingly, in this instance, the firm uses computer technology to meet requisite variety; it maintains a database that provides traceability of materials that would otherwise be impracticable.

Quite different arrangements of personnel may be involved in achieving regulation; for example, teamwork in problem solving. In the dental laboratory, certain new disturbances (for example, when established work methods would not work with new materials) were solved through team work where the apparent two-layer hierarchy and technical segments would merge into a flat team to generate a solution.

2.4.3 The R_0 Level - perspectives on the operative/craftsperson

In much of industry, whether by virtue of downsizing, enlightened views of job enrichment or pragmatic necessity, operatives are increasingly granted degrees of autonomy in their work. In larger firms the decision-making aspect of work had been somewhat obscured from managerial consideration by Tayloristic assumptions in the design of work. Within the Tayloristic perspective, the worker has been valued more for their hands and ability to perform robot like repetitions–in essence, much more T than R. Responses to D would tend to be performed according to set programmed responses handed down from R_1, rather than generated by R_0.

In small firms, the potential of operatives to function as true regulators is more likely to be exploited. The case studies provide examples of operatives being expected to sort out problems of an operational nature. For example in the bakery, the pastry baker recounted episodes of investigation into baking problems involving experimentation with flours of different gluten contents, water content etc. Similarly, within the dental laboratory, technicians developed solutions to problems within their own specialisms.

This example demonstrates a principle of regulation much exploited by small firms. In order to achieve a state close to requisite variety, R_1 relies on the resources of R_0 to adapt to changes in D. However it is also clear that in order to maintain control, R_1 retains authority over the questions of which essential variables are measured, and assessment of whether the values are acceptable or not.

2.4.4 Perspectives on Maintaining control - What is important, what is wanted

The essential variables E, represent what is important, typically the performance indicators used by managers in small firms to monitor progress. The η set, by contrast, represents the set of standards that will be used to assess whether any of the essential variables represent unacceptable performance. The essential variables E are what is important, the η set, what is wanted.

The evidence of the case studies suggests that E and η are defined differently for quality issues, as against health and safety issues. Health and safety does not appear to be subject to the same feedforward and feedback communication as production essential variables which affect quality. In two of the three firms we looked at, the R_1 attitude can be summed-up as "well I've told them, I've given them PPE, it's up to them". In other words, R_1 is delegating responsibility to R_0; it is up to R_0 to decide whether it is acceptable or not to wear PPE. To put it another way R_1 has decided PPE is 'what is important', but leaves it to R_0 to decide whether it is wanted'. In contrast, R_1 would certainly intervene if for example R_0 failed to maintain lengths of steel cut within set tolerances.

The logic of requisite variety suggests that R_1 will preserve as much freedom as possible for R_0 to make decisions, so long as the η set values are not infringed. The less freedom R_0 has, the less amplification R_0 can achieve. Health and safety is perceived by many people as equipment based add-ons that reduce the freedom of operatives. So a production-focused R_1 might see the use of PPE as a reduction of operating freedoms that might jeopardise production, whereas a health and safety focused R_1 will accept the reduction of operational variety as a small price to pay to safeguard health and safety.

2.4.5 Perspectives on transformation and adaptation

Our R_0 welder is not told how to weld by R_1, nor does R_1 tell our R_0 baker how to bake. What to weld, what to bake are choices made by R_1, these represent in themselves massive incursions into the operational variety open to R_0. The final choice of one from many, is left to the operative (this is how R_0 amplifies R_1). Part of this residual variety is the process of adapting responses to hazards inherent in the operative's tasks. This process of worker adaptation has its place in risk management, but its effectiveness depends on various assumptions that are not necessarily apparent to the higher regulators who rely on adaptation to deliver healthy and safe work.

The engineering firm we studied had recently suffered an accident, a serious event involving a deep cut to the victim's upper arm from a circular saw. Talking with fellow operatives produced some interesting comments, "It wouldn't happen to one of us", 'well he was new here". The assumption amongst the operators seemed to be that they were fully adapted, simply by experience, to the set of disturbances hosted by the business. This mode of establishing requisite variety in a regulator, learning by doing, is very powerful, and widespread. One would expect this approach to be used when D and/or T are complex and the channel for communication between R_1 and R_0 is limited (the capacity being determined by various characteristics, not least by time but also technical vocabulary). One would expect the limitations of R_0 to R_1 channel capacity to be particularly evident when new or inexperienced operatives staff the R_0 function. Due to limited feedback capacity R_1 is very unlikely to be fully conversant with the range of D which R_0 may meet, and almost certainly would be unable to transmit to R_0 sufficient of the full range of skills and values required in the time available.

2.4.6 Perspectives on Skill-, Rule- and knowledge-based behaviour

The skills-, rules- and knowledge-based scheme of behaviours developed by Rasmussen is complementary to and provides additional insight to the DTER regulatory framework. Skill- and rule-based behaviours are appropriate to states of DTE that are relatively frequent. Skill- and rule-based responses require less data processing capacity from the regulator; they are effectively pre-programmed responses. Constant repetition tends to ensure that they remain effective tools to maintain the set of essential variables within acceptable tolerances.

Skill-based	"Automatic" behaviour governed by memorised sequences of actions, reinforced through frequent application
Rule-based	Behaviour characterised by correcting familiar problems by the application of previously established rules of an if X (state) then Y (remedial action)
Knowledge-based	Behaviour determined by conscious analytical processes occurring in real-time

Table 2 Description of SRK levels of behaviour

For less frequent states of DTE that may threaten the η set, some other means of achieving appropriate regulatory response is required. Within Rasmussen's SRK scheme, this means knowledge based behaviours. This implies a more active cognitive process (the discussion of responsibility at section 1.3, page 4 is germane here). In simple terms, this will involve a search for information to identify and analyse the problem, a comparison process of the information found against that provided by the regulator's mental model of DTE, and a selection of responses to deal with the problem. In other words, a more abstract understanding of DTE behaviours that allow the regulator to develop a response that will block the disturbance of the η set. For example, the bakery flour silo would occasionally be overfilled or filled using too much pressure (D). This would cause various problems (departures from η such as spoilage of flour, interruption of production to clean-up etc). The development of a regulatory response to this involved analysis of the problem to identify the various points of intervention, including acting upon the source of the disturbance itself (the flour mill). However, R_1's abstract knowledge remains deficient in this matter (that is below <u>requisite</u> variety) because despite various attempts by Ms May (and previous HSE contact on the same matter) R_1 did not appreciate the possibility of dust explosion.

2.4.7 *Perspectives on responsibility at the R_0 level*

Interestingly, operator understanding of the processes they control is often a product of small variations of conditions or methods as well as more abstract information received through training and education. In all of the small firms we visited, the role of training and education was evident and important; in the case of the engineering firm and the bakery, lengthy apprenticeships, in the case of the dental laboratory day-release higher qualifications.

The upshot of this is perhaps far more important than it first appears: whether rightly or wrongly, the regulation of health and safety of people in small firms is achieved in large part by amplifying the individual variety of the owner/manager. Where the disturbances are of a type, inherently or by effect, that are not amenable to adaptation-through-error in the workplace, the regulatory variety of the firm needs to be supplemented externally.

Another situation, where adaptation-through-error is unlikely to result in requisite variety, arises where R_1 does not insist upon the values that define the η set for R_0 – the circumstances of "routine violation". Even η set norms established by previous employment or external training are likely to become weakened if they are not reinforced as part of regular experience. Such norms may also be actively challenged by reinforcement of other essential variables that are at odds with the norms of the individual.

2.4.8 *Perspectives on Adaptation*

Some disturbances may be beyond the scope of R_0 and will require an R_n intervention to provide adequate regulatory response. Sometimes this might correspond to changing T_0 in some radical way or influencing the source of D beyond R_0's realm of influence. In the bakery we visited, an R_1 response to the overfilling of the bakery's flour silo was to discuss the matter with the flour mill to get the maximum size of delivery reduced. However, given that in some work systems in very stable market environments (i.e. the set of D does not change), the habit of adaptation (and the maintenance of channels to information that underpin adaptation) at the Rn level is unlikely to be sustained as novel challenges to the η set are rare. This appeared to be the case in the engineering firm we visited; a very stable market niche with little change to the associated technology in 20 years. Information channel bandwidth may be available to this company (via the exchange of goods and services) but the messages are low variety and connected with production aspects of η. Outside of highly stereotyped communication, this firm was in an information "bubble" disconnected as it were from the "environment of compliance" of essential variables such as health and safety.

Furthermore there are all manner of psychological devices to reduce the significance of failures (i.e. when one or more essential variables, E, fall below or exceed the acceptable range, η). Aside from the host of emotional and cognitive explanations as to why people try to avoid changing, a simple signal-detection paradigm predicts (on the basis of abundant evidence from industrial studies of vigilance, Wickens, 1992) that people require a high standard of evidence before embarking on an expensive process of change (in whatever units of expense are appropriate). Thus even where there are small unwanted events that serve to warn of future, related but more serious events these may not trigger a problem-solving cycle (itself a method of replenishing missing regulatory variety). This need not be seen as irrational: as Pineda et al (1998) suggest, it is boundedly-rational in the Simon (1968) sense of being consistent with a less-than-requisite regulatory. On occasions where falsification threatens, the evidence can be explained-away using a host of schemes ranging from outward blame to fatalism and other superstitions.

2.4.8.1 Examples of poor adaptation

In the engineering firm that experienced the accident described above, R_1 was a hard pressed director/operator who divided his time between co-ordinating the operations, providing extra operational support for "rush" jobs, acting as a general purpose machinist (currently doing the work of the IP) and also other duties including overseeing the newly employed. There were various options open to him to exploit the constraint evident in D. That is to recognise that not all operations were equally dangerous, or equally probable and to apply scarce resources to those which were both dangerous and probable, and to use the strategies mentioned above. However, the presiding beliefs were that things were safe – based, not unreasonably, on a low accident frequency. Of course, a low accident frequency can be attributed to a tame environment or, as in this case, a ferocious one that is populated by skilful "old hands" who have learned where to put their fingers, when to look away from welding flash and soon. In the case of R_1, the hazards of his workplace had become wallpaper, providing no information to trigger appropriate hazard controlling responses.

Despite the accident, touring around the worksite with R_1 it was clear that the variety he demonstrated as a regulator was greatly limited in comparison to the variety demonstrated in the hazards detectable in the workplace. The less obvious hazards, those that threatened health, were particularly unlikely to be recognised. Even Ms May explained the hazardous aspects of his workplace; R_1 was unclear about what to do.

Talking with some of the R_0 operators in the engineering firm, three issues impressed themselves. First, the importance of apprenticeship in developing the skills and knowledge of the craftspeople and the sense of pride and identity it seemed to confer. Secondly, there was an almost total absence of external training since serving apprenticeship. Thirdly, the low variety, that is, the slow-changing and stereotypical, nature of the workplace risks to health and safety. The health and safety of the workforce at the engineering firm depended on the assumption of no change; no change of plant, people or process. The standards demonstrated in the plant were those accepted by the system of regulation of the firm, they may have met the standards of 10-15 years ago but they did not fully meet the current standards laid down by HSE.

The health effects of welding fumes provided a specific example. The staff judged visibility to signify the D variable and triggered a response to control the fume. This provides some error-controlled regulation - but not necessarily enough to prevent D becoming a hazard to health.

Also, perhaps more seriously, the firm does not appear to have much flexibility to adapt to external changes in D. A new powder paint treatment, or an unusual combination of orders, oversized materials etc., might easily stress the fragile balance of D, T and R. Effectively this was the case with the newly hired victim of the accident they had recently suffered.

> It is worth noting in passing that the case studies reported in this study suggest that the flexibility so often asserted of small firms may be a feature of the whole *sector* that is not necessarily replicated in the *part*. As Lybaert (1998) observes, the flexibility exists in the *ecology* of small firms; in the richness of niches exploited, and not necessarily in the ability of each of these firms to fulfil other niches made available by changes in the marketplace (so-called "niche-dynamism"; Dean et al, 1998).

2.4.8.2 Conditions for successful learning

An important question is how in general does the regulator become adapted? In many cases, and this is true at any level of R, adapting behaviour (RT) in response to small variations around the margins of the eta set provides an important mode of learning. Two conditions must be met for this mode of learning to be effective:

(a) The trainee must recognise the links between disturbances D, and variations in the eta set values
(b) The trainee must be suitably averse to essential variables E moving out of the eta set range of acceptable values.

In small firms, this mode of learning is likely to be more relied upon than in larger firms. In smaller firms there is likely to be less reliance on learning outside the workplace (true of the engineering firm and bakery, decidedly not true of the dental laboratory). However, in smaller firms R_1 can provide safer learning conditions by:

1) pointing-out and reinforcing the connection between states of D and state of the eta set
2) closely supervising R_0 in hazardous operations to indicate to R_0 states of D, T and the eta set
3) indicating across a range of operating parameters, appropriate control actions on T until R_0 is experienced enough to work unsupervised

2.4.9 Perceptions on Supplementing variety

2.4.9.1 Life-cycle and other triggers to supplementation

The feedforward route, which is more-or-less coincident with "proaction" in H&S, will be triggered at different times and to different extents depending upon business size and type. In our case studies, the clearest example of an episode of feedforward was the start-up phase of the dental laboratory. Here, the owner had gone to great lengths not only to plan but also to actively seek the participation of external sources including regulators such as HSE, environmental health etc. Certainly, the "start-up" phase is a clear opportunity for external regulators to increase the variety of self-regulation of firms. Banks attempt this by determining supplementation of the firm's resources around a business plan, TEC's attempt this by offering (free in many areas) a 2 to 4 day "enterprise" training course which provides regulation and supplementation of the firms start-up planning.

It may well be that there are other lifecycle phases characterised by active acquisition of information to improve self-regulation. For example: expansion, new product development, acquisitions etc. What is clear from the literature is that small business sector <u>has internal variations</u> in this respect. For example, Agarwal (1998) suggests the type of technology appears to be a discriminating factor (as the dental laboratory and engineering firms illustrate). Several studies (Matthews and Scott, 1985, provide a useful review) acknowledge that entrepreneurial firms are different from other small businesses in the means, quality and quantity of their information handling.

The trigger to a feed-forward supplementation might well be a signal that requisite variety is not being achieved (such as unforeseen incidents) or a change in η (in the conservative direction) or a change in the components or structure of the operational system, T. To this extent, accidents and near-misses can be low-variety messages (analogous to acute pain) that may signal the need for high variety replenishment of regulatory knowledge to regain requisite variety.

2.4.9.2 The Involvement of Network in Supplementation

As mentioned earlier, the dental laboratory provides a clear example of a firm that has developed large (information) channel capacity to sources of supplementation. Their marketplace is characterised by rapid change and adaptation requires "running to keep-up". Interestingly the dental laboratory used all of its employees to this end: the day-release trainees provided a channel for exploiting the resources of the colleges. Similarly, regularly attending seminars, taking training courses and reading trade journals were functions undertaken by everyone in the firm. Interestingly, the operational organisation of the firm showed evidence that its director had attempted to involve the local TEC, Chamber of commerce in his business to similar systemic ends.

What did not come through the case studies but which is strongly apparent in the literature is the importance of network and alliances to the success of small firms. The general belief is captured by Yli-Renko and Autio (1998):

> "according to this view, the key to organisational survival is the ability to acquire and maintain resources. Firms enter into network relationships with other organisations because they cannot generate all the necessary resources internally. This interaction results in high interdependence of organisations, as they form relationships with those who control external complementary assets. Through this interdependence, organisations are embedded in an environment comprised of other organisations"

The view of network resources extends beyond other firms to the social linkages to the small business. Gulati (1998) provides an extensive review of the literature on this wide interpretation of network operation.

External assistance of networks can be seen in terms of regulatory design – or, in Aulin's terms, increasing the self-steering capacity of industry at the level of the sector rather than the level of the firm. Rabellotti (1998) gives examples of UN programmes aimed at providing this assistance in developing economies. Lastly, the notion of network operation as a business competence in its own right, as proposed by Zajac, 1998 has strong parallels to the policy orientation in the European Commission (EC, 1996 and 1997). In the latter case, the term *informacy* is used to describe the whole raft of skills and knowledge that allow business people (and individuals in society generally) to properly exploit sources of supplementation.

2.5 Summary Observations from the Case Studies

In all cases, the health and safety standards of the firms examined were less than adequate. This section has considered some of the reasons why this might be in terms of the flows of information argued to be necessary to successful regulation. Although the purposes of the case studies was to provide illustration of the cybernetics ideas under discussion, various themes did emerge that may generalise beyond the three firms visited.

- The self-regulation of health and safety essential variables may be subject to different handling than those of production. These differences may reveal points of intervention for compliance activities.

- Learning through error is a powerful aspect of self-regulation but not a method that necessarily favours health and safety essential variables. Methods to encourage risk assessment in the small business sector need to recognise the discontinuity between the feedforward mode assumed HSE publications such as "Five Easy Steps" and the dominance of error feedback.

- From the human factors perspective, errors at the knowledge-based level of Rasmussen's SRK model holds particular threat in small business such as the bakery and engineering firm as the supplementation of abstract knowledge is very limited.

- The dental laboratory illustrates how continuing vocational education provides bandwidth to HSE's channel of supplementation. In addition, opportunities to support small businesses through social and business networks, intermediaries and ICT (Information & Communication Technology) are open to development by HSE in their role as designers of the self-regulatory health and safety system.

- The leverage that outside agencies have on small businesses may vary with the firms' lifecycle. Firms may be more receptive to regulatory input at start-up than after. Different phases of the lifecycle may call for different contact strategies.

3 IMPLICATIONS FOR SELF-REGULATION

It is hoped that the foregoing sections will have provided clarification of the principles under discussion. This section discusses how these principles could be applied to HSE's task of improving standards of health and safety protection in small firms: broadly, the fostering of self-regulation in the small business sector.

It should be recognised that much of what follows already exists in some state of development, that is, the suggestions are not radical departures from current practice. The application of these principles to HSE's work in the small business sector is seen as an aid to planning and, in particular, a means of reconciling different HSE initiatives, roles within a single framework in which they can be ordered and discussed. This facilitating effect is captured by Espejo & Harnden (1989):

> "A model is expected to provide a setting, a common frame—in other words, it is expected to make visible a set of constraints, within which certain problems can be enunciated in a particular way, and certain problems solved.
>
> Let us be clear about this. A model is a convention—a way of talking about something in a manner that is understandable and useful in a community of observers. It is not a description of reality, but a tool in terms of which a group of observers in a society handle the reality they find themselves interacting with. … But whatsoever, an individual may never communicate what is accessed to another individual except in terms of models. This is not a limitation, but is precisely the motor for the generation of a consensual domain. A consensual domain is none other than the play of a particular set of interacting models."
>
> Espejo and Harden, 1989, Page 446

As may already be apparent, HSE itself can be considered within the picture of the regulation of health & safety. A systemic viewpoint may benefit from a big-picture interpretation of the working through of the three senses of responsibility in Society. There is also benefit to be gained by reconciling the various visions of the work of HSE/C. Reconcile does not mean standardise but it does emphasise *discussion within a framework oriented to a purpose*, in this case the ambitions of society and government.

3.1 Fostering Self-regulation in the Small Business sector

The upshot of the Ashby-Conant maxim (discussed on page 6) is that regulators have no choice but to operate through a model of the system they regulate. All regulation is determined by actions based on how the real world is reflected in the regulatory model.

Given the importance in regulation, the establishment and maintenance of up-to-date regulatory models is clearly a matter that deserves explicit treatment. However, rather than a single model, it is most likely that a number of models would exist within a guiding-set of conventions. It has been argued that the cybernetics literature provides an explicit set of conventions (ultimately reducible to mathematical form) that would serve this purpose. The medium of these models is a matter best determined by the purposes of their users/owners. Some may remain as mental models manifested through action or conversation; as computer simulations; as diagrams; as the data structures underlying statistical work.

The topic of how to establish models has been introduced in section 1.4, page 6. There the main concern was that the behaviour of the model should be sufficiently close to that of the real-world to interpret data arising from it and to guide regulatory actions to achieve the desired regulatory outcomes. In the present context, a model of the UK small business sector appears to be a pre-requisite to HSE regulation at the strategic level. As with any modelling, the supervening issue is the purpose that the model serves. Which aspects of the real world to preserve in the model, and in what detail, can only be decided by appeal to the purposes of the system. To decide what to include within a *regulatory* model, we need first to observe Ashby's prerequisites – "*Before any regulation can be undertaken or even discussed, we must know what is important [as outcomes] and what is wanted*". If we circumvent this step there is no safeguard from models that lack variety because no *real* criteria of success have been established. However, it is possible to begin a model on a scant empirical basis as the seed around which increasing structure can be added as information is established through experience. The knowledge we lack can be gained from the system regulated, as a consequence of regulatory activity; as it was put earlier: *it is mice that teach kittens how to catch mice*. However, this cannot work in the absence of clear aims for regulation and, clearly, aims that are matched by valid operational data concerning behaviours in the regulated system. In general terms, the model needs only to be amenable to data that make a difference to regulatory choices at the level in question. Data that do not make a difference to regulatory actions ($R_n \rightarrow T_n$) need not be collected.

In addition to operational data, Beer (1985) suggests that two further classes of information (and corresponding supplementary channels) must be established to challenge the model and cause it to be reformed. This need arises because regulatory models in use tend to be, in the end, subject to increasingly denatured channels of feedback. There is a variety of reasons why this occurs, not least that a successful regulatory scheme tends to reduce the amount of information arising through feedback. In many systems, including social systems, communication channels tend to "lose tone"–dena-ture– if not used. Furthermore, in social systems, there may be other factors that tend to attenuate feedback and allow models to persist despite the lack of variety that should cause them to be reconsidered. Whatever the reasons, a regulatory model that is prey to such problems will not be able to maintain requisite variety. This is true at every level where an adaptive mode of regulation is used, including those regulators beyond the HSE boundary.

In advance of debate about the aims for regulation, we can reflect on some of the likely features of the "seed" model. It should accommodate the regulators (within and outside of HSE) that amplify government policy (i.e. the aims for regulation) into actual practices at the operational level in industry. This should be as open-minded as possible as to what counts as a regulator; certainly including intermediaries such as industrial bodies, trade associations and Business Links but also colleges, families, business-networks. Similarly, the model should identify likely channels that connect the various levels of regulation with the operational level. At the lower levels of the model, we could treat, until experience teaches otherwise, that all small firms are replicates; that each firm has certain connections with intermediaries (that are themselves replicates) and so forth.

3.2 Regulatory models of the Small Business

Here, the modelling may be chiefly concerned with the small firm as a self-regulatory system with connections to its environment including sources of supplementation and inputs from higher-order regulators. The significance of this is that practical work in the area health and safety tends to rely on the paradigm of health and safety management; this may need modification to accommodate pertinent attributes of self-regulation not present or adequately emphasised.

As previously argued, the level of regulation concerned will determine the form of the model. At the level of the inspector, a low-variety model of small firm self-regulation that is amenable to amplification in the particular case could act as a guide to inspection and as diagnostic tool. For example, the low variety (i.e. rudimentary) method used in our case studies acted as a guide to developing a higher variety model of regulation in each small firm. This understanding when coupled to the state of compliance (meaning, in this case, the standards of health and safety in the workplace and work practices) provides a basis for diagnosis.

The next issue for the low-variety model is to associate diagnosis with treatment. For example, the situation of the engineering firm in our case study was a chronic one. Remedy of the specific hazardous conditions and practices would itself count as one level of treatment (palliative, as it were). The next level of treatment relies on understanding – diagnosing – the properties of the regulatory situation that allow this situation to persist. In this case,

1. the firm has little or no connection to sources of supplementation (external current training or other forms of instruction for personnel whether operative of supervisory);
2. the channel capacity connecting the higher order regulatory hierarchy with the firm has insufficient bandwidth (at least as regards the channel capacity relevant to health and safety essential variables);
3. R_1's model does not have requisite variety with regard to Health and Safety essential variables
4. The predominant regulatory mode of error-controlled feedback does not allow adaptation to forms of D and harm to E beyond the sensory range of R (therefore Health issues are particularly at risk)

<u>Some</u> of this can be addressed at the inspector level: improvement notices for training might help at (1) as may contact with an appropriate intermediary organisation (which may also have a role in treating (2). Causing R_1 to recognise the inadequacy of his regulatory model might be enough to initiate R_1 to allocate time to remedy (3 and 4).

As well as providing a basis for interventions in the field, the same model would provide a scheme for data to be transmitted upward in HSE. This upward transmission would allow higher level regulation to occur (i.e. at the level of the sector, geographical area, industry etc). This is likely to have two forms – one form aimed at the individual instance, the other focussing on the class of which the dutyholder is an instance. Concerning the first, this is amplifying the inspector (e.g. back-up for the inspector contact by allocating WCO support, placing the dutyholder on targeted mailing lists for leaflets, discussion between HSE and the appropriate intermediary groups etc). The second form is the use of the operational data relayed by the inspector as strategic intelligence to inform higher-order regulation (as alluded to in the penultimate paragraph of section 0).

The foregoing emphasises the requirement that different levels of model be compatible. In practice, there are likely to be different models at different levels, each having requisite variety. Put simply, the models developed in real-time by inspectors in dutyholder premises will have more <u>detail</u> than the HSE strategic equivalent but both should strive to have requisite variety.

3.3 HSE as a designer of regulators

Earlier, it was stated that design and regulation are as equivalent within cybernetics. There is clearly an argument that HSE could focus on developing self-regulatory capacity of whole industries or sectors of employment. This already happens but there is room to further inter-connect or "vertically join-up" way (to borrow the slogan of "joined-up thinking in government").

The emphasis here is on design as regulation; just as R_1 has a role in the design of R_0, so HSE has a role in the design of R_1 or any level of regulator. This argument includes the design of regulatory levels of regulation within HSE's organisational boundary. However, what is distinctive about this is that it is separable from the paradigm of health and safety management as espoused in publications such as "Successful Health and Safety Management". This allows the *de-coupling* of the health and safety *message* (i.e. promoting the health and safety component in η and the technology of D_0 and T_0) from the *medium* that has a more lasting influence in the workplace (management). There are many advantages. First, it allows common goals to be easily identified and discussed between different regulatory parties (e.g. different government agencies, between any given government agency and intermediary body). Second, it permits the development of common models and measures of self-regulation that can be mutually serviced in industry by all regulatory agents. Third, it encourages identification of those aspects of health and safety that truly require exceptional treatment.

The upshot for HSE activity is to how to understand and remedy problems in businesses that fail to deliver a η set characterised by SFARP (which may extend beyond the boundary of a particular duty-holder to its industrial milieu). This may well be facilitated a diagnostic approach that is informed about health and safety issues but not dominated by them.

At the strategic level, the ideas presented here may facilitate consideration of regulatory structures (for example intermediaries "between" government and employees) that do not currently exist or, where they do, have scope for improvement. To paraphrase Ashby, Cybernetics takes as its subject matter the domain of "all possible machines" and is only secondarily interested if informed that some of them have not yet been made, either by Man or by Nature. What cybernetics offers is the framework on which all individual machines may be ordered, related and understood".

However, the means by which modellers apply systemic principles will vary with their needs. For example, methods such as Beer's Viable System Model may facilitate modelling where DTER becomes clumsy and vice versa. Models of the real world rendered through DETR or VSM can correspond because they are projections of the same underlying principles.

Inspector-led regulatory design

Many channels mediate the relationship between HSE and dutyholders; Inspectors to premises provide a vital channel. Ironically, whilst this channel is regarded as very important by the inspectors themselves and in the public imagination, the bandwidth it provides between the small business sector and HSE is limited. Limited though it is, inspection has potentials not easily afforded by the other communication channels at HSE's disposal. The advantages include

- the preference of small business people for personal (as against impersonal) sources of information;
- the opportunity for the inspector to gain information (in the sense discussed on page 3) about the business;
- the opportunity for the inspector to alert the business to sources of information (e.g. intermediaries, HSE helplines, internet resources, other small businesses (if appropriate) texts

To gain these advantages requires a set of conventions that enable the inspector to undertake assessments of this type, the organisational back-up to amplify the actions the follow from this assessment and an efficient interface between the inspector and the organisation to facilitate this.

The phrase "set of conventions" is used with Espejo's connotation: both a model and method of modelling: as much a means of discussing as a means of analysis.

On the inspector assessment side, the Ms May application of the method used in the case studies demonstrates that the basic approach is useful and feasible. Further, that it adds value to inspections with little cost in time and effort. However, the case-study method was designed only to furnish this report with illustration of the principles involved. As mentioned in section 4, the development of a tool for inspectors would need explicit treatment to ensure usability and effectiveness.

Lastly, "organisational back-up" as mentioned above would want to go beyond assisting the one firm identified by the inspector but would aim to generalise the inspectors assessments, where appropriate, to the targeting of interventions in comparable firms. This is discussed further in section 4.

3.4 Amplifying HSE

3.4.1 Designing intermediary regulation

What we know about small firms through the literature and, what we expect by virtue of the current approach, suggests that we need to reflect upon the adage that "good health and safety is good business". The philosophy of the slogan is commendable but may not always mesh with the underlying pattern of actual business practice.

For example, risk assessment, however simplified, is a feedforward ($D_0 \rightarrow R_n \rightarrow T_0$) activity that does not sit well with the highly economic (from the perspective of R_n resources) error-controlled option ($D_0 \rightarrow T_0 \rightarrow E_0 \rightarrow R_0 \rightarrow T_0$). We might argue that the firm would generally benefit (i.e. across all classes of essential variable) by adopting a more consistent approach to planning. Indeed, Rue and Ibrahim (1998) provide a review suggesting there is empirical support for planning as beneficial in itself to small business). However, taking account of the wider context of business practices and reasons for their use may well advance the aim of achieving change in small business health and safety practices.

In systems where low-level regulators lack variety (a low degree of self-steering), Aulin suggests that that regulation can be increased by adding more levels of hierarchy until the desired level of control is reached. This opens possibilities but also raises the spectre of the old "prescriptive" approach to health and safety regulation. However, in the light of a quarter-century of rapid change, the utility of a prescriptive approach may require re-evaluation. In particular, ICT provides bandwidth that did not exist previously. As noted in the 1996 EC Green paper *"Living and Working in the Information Society: People First":*

> *"ICTs also offer new opportunities in the field of health and safety at work. In particular, the application of new technologies can bring considerable added value to risk assessment activities, the collection, screening and dissemination of information, education and training in occupational safety and health, and for end users, particularly SMEs."*

Additionally, social structures can provide intermediate stages of regulation that achieve amplification of HSE variety. This may provide "prescriptive" options but achieve requisite variety (subject to appropriate style of interaction) perceived by SMEs as facilitation rather than limitation and burden.

Requisite hierarchy: self-selection by small business

ICT and efficient intermediary networks also offer a change of focus from the "paternalistic" view of small businesses as needing intervention and guidance from, as it were, above. The vision here is optimisation of the currently patchy operation of hierarchies that support only the relative few that demand attention by highly visible difficulties. Often the "support" in health and safety terms is a momentary high variety "burst" often focused at T_0 rather than the regulators of the firm (from R_0 to R_n). Whilst the actual specifications are moot, the vision is of a system whereby dutyholders essentially select a level of "being told what to do and how to do it" on the understanding (a) they contribute the necessary funds (for training, equipment, maintenance etc), and (b) that the reciprocal accountability arrangements are correspondingly weighty. This is analogous to the resource bargain (a term coined by Beer, 1989) as might exist between the division and corporate centre of a large firm or between a start-up firm and a bank. The firm trades certain freedoms of choice in return for short-term solution of its current problems and longer term commitment to developing its own competencies.

The important prospect is that the firm becomes increasingly able to self-steer, after the initial investment of variety and period of actual operation, stable. As the regulatory capacity of the firm becomes requisite to the variety entailed by its operations, the firm can reduce the height of hierarchy above itself. This view whilst seemingly utopian is not necessarily so; effective industrial partnerships demonstrate much of the characteristics outlined. Similarly, it is largely communication channel capacity between the state and small businesses that requires a "one-size-fits-all" approach to enforcement, effective use of intermediaries and ICT may provide bandwidth for a much more flexible and tailored approach.

Mutual absorption of variety

The law of requisite variety can be stated as "*only variety destroys variety*" and this report has discussed many of the options for increasing regulatory variety. However, a frequently available source of variety is available at the same level that requires regulation. Simply, one business can help absorb some of the variety of another business. The rational behind this is that complex systems tend to exhibit self-organising capabilities, even when the data exchanged in internal relationships of interest appears to be somewhat undirected or random (Klir, 1991). In addition, firms that recognise their connection to a shared source of disturbance can take advantage of a mutual supplementation of regulatory variety. Supply networks are another case, although different as sometimes the focus may be to reduce a *transfer of disturbance* between firms in supply relationships.

FOD already uses this idea, for example, exploiting networking by encouragement through seminars. What is emphasised here is that there are many opportunities where this effect can be harnessed to achieve HSE ends. What is needed is the skill of recognising when these opportunities are available. For example, HSE may be able to take advantage of existing business networks and joint-ventures to facilitate co-operation between firms on health and safety issues or making connections for firms identified as relatively isolated.

4 FURTHER WORK

4.1.1 Developing a taxonomy of small businesses

The modelling work discussed in sections 0 and 3.2 could be elaborated in two ways. First, to further develop the case-study method as a means for gathering data about the distinctive modes of self-regulatory failure. The goal of this research would be to establish a diagnostic system amenable to association with treatment (or as a trigger to further R&D to reveal the underlying reasons and development of "treatments").

The second strand of research would take this work beyond the provisional view of small business as replicates. This would aim to develop taxonomy of small businesses using their informational characteristics. This would facilitate recognition of firms and guide HSE's interaction with them in a way that is efficient in resources and sensitive to the needs of the firm self-regulatory capacity.

4.1.2 Developing an assessment tool for the inspection toolkit

The two strands of research mentioned both support the development of an assessment tool for use in the field. This would have several requirements, not least of which is usability for inspectors. In outline this would aim to assess the type of firm (within the taxonomy) and have a diagnostic function to identify problematic areas of information handling that require change to achieve appropriate health and safety performance. The identification of firm type and problem will lead to choice of action to improve the firm's self-regulatory capacities.

As a history of contacts with firms is established, the associations between the taxonomy, diagnoses and interventions can accrue. This will provide the model of small businesses with a learning facility to which knowledge management processes (e.g. as set out in van der Spek, 1997) can be applied.

4.1.3 Linking taxonomies with interventions, implications for Best Mix

As mentioned above, the aim of this work would be to facilitate recognition of firms and thus guide HSE's interaction with them in a way that is

- efficient in HSE resources
- sensitive to the self-regulatory needs of the firm

Just as the inspection tool is supported by a learning model that provides low-variety direction to an inspection, FOD's "Best Mix" strategy will be given requisite variety by a learning model that associates the type of firm with a contact method. The feedback verification of the impact of the contact method will provide Best-mix with an adaptive mechanism.

This approach also suggests the means of extending Best Mix to new forms of exploitation. As discussed earlier, a model that "recognises" a firm by its information handling characteristics and has requisite variety of response may provide firms with a means of self-help (e.g. using ICT to provide the channel bandwidth). Closer in time, the approach also suggests that FOD can further support the efforts of field inspectors with back-up responses to low variety messages. For example, if HSE were to have a well developed taxonomy, a discriminating diagnostic set and a requisite variety of actions (e.g. variety of leaflets, contacts with intermediaries, help-lines, websites etc.) simple triggers will be enough to initiate centrally supported intervention (or cascade of interventions) with the firm in question.

5 References

Agarwal, R. (1998). Small Firm Survival and Technological Activity. Small Business Economics, Vol. 11, pp. 215-224.

Ashby, W.R. (1956). Introduction to cybernetics. London, Chapman and Hall.

Ashby, W.R. (1960). Design for a brain. 2nd edition, London, Chapman and Hall.

Atlan, H. (1983). Information theory. In: Cybernetics Theory and Applications. Edited by Trappl, R., London, Hemisphere(1983).

Aulin, A. (1982). The systemic laws of social progress, towards a critical social philosophy and a criticism of Marxism. Oxford, Pergamon Press.

Aulin, A. (1989). Foundations of mathematical system dynamics, the fundamental theory of causal recursion and its application to social science and economics. Oxford, Pergamon Press.

Beer, S. (1985). Diagnosing the system for organisations. John Wiley & Sons, Chichester.

Beer, S. (1989). The viable system model: its provenance, development, methodology and pathology. In: The viable system model: interpretations and applications of Stafford-Beer's VSM. Edited by Espejo. R., and Harnden. R., John Wiley & Sons, Chichester. pp. 39-50.

Conant, R. and Ashby, W.R. (1970). Every good regulator of a system must be a model of that system. International Journal of Cybernetics, 1970, Vol. 1, No. 2, pp. 89-97.

Dean, T.J., Brown, R.L., and Bamford, C.E. (1998). The differences in large and small firm responses to environmental context: strategic implications from a comparative analysis of business formations. Strategic management Journal, Vol. 19, pp. 709-728.

Espejo, R., and Harnden, R.J. (1989). The VSM: an ongoing conversation. In: The viable system model: interpretations and applications of Stafford-Beer's VSM. Edited by Espejo. R., and Harnden. R., John Wiley & Sons, Chichester. pp. 39-50.

European Commission (1996). Green Paper: Living and Working in the Information Society: People First. COM (96) 389 final. Available via URL: http://www.ispo.cec.be/files/peopl1st.doc

European Commission (1997). Final policy report of the high-level expert group: Building the European Information Society for us all. Unit V/B/4 Directorate-General for employment, industrial relations and social affairs. Available via URL: http://www.ispo.cec.be/hleg/hleg.html

Fitzgerald, R.M. (1996). Matched-phase noise-reduction. Journal Of The Acoustical Society Of America, 1996, Vol. 99, No. 3.

Gulati, R. (1998). Alliances and Networks. .Strategic management Journal, Vol. 19, pp. 293-317.

Heylighen, F. (1997) Classic publications on complex, evolving systems: a citation-based survey. PO, Free University of Brussels, Pleinlaan 2, B-1050 Brussels, Belgium. Available at URL: http://pespmc1.vub.ac.be/Papers/ PublicationsComplexity.html

HMSO (1972), Committee on Safety and Health at Work Safety and Health at Work: Report of the Committee (Robens Committee), London.

Kingston-Howlett, J.C. (1996) The Evolution and Design of Safety Management Systems. Unpublished PhD Manuscript, Aston University.

Klir, G.J. (1991). Facets of cybernetics. Plenum Press, New York.

Levi, M. and Powell, P (1998). SME Flexibility and the Role of Information Systems. Small Business Economics, Vol 11, pp 183-196.

Lybaert, N. (1998). The information use in a SME: its importance and some elements of influence. Small Business Economics. 10: 179-191.

Matthews, C.H., and Scott, S.G. (1995). Uncertainty and Planning in Small and Entrepreneurial Firms: An Empirical Assessment. Journal of Small Business Management, October, 1995.

Pineda, R.C., Lerner, L.D., Miller, M.C. and Phillips, S.J. (1998). An investigation of factors affecting the information-search activities of small business managers. Journal of Small Business Management. Vol. 36, No.1, pp. 60-71.

Rabellotti, R. (1998). Helping small firms to network - the experience of UNIDO. Small enterprise development, Vol. 9, No. 1, pp. 25-33.

Rasmussen, J. (1996). Risk management in a Dynamic Society: A Modelling Problem. Key-note address: Conference on human interaction with complex systems. Dayton, Ohio, August 1996.

Rue, L.W. and Ibrahim, N.A. (1998). The relationship between planning sophistication and performance in small businesses. Journal of Small Business Management., Vol. 36, No.4, pp. 24-32.

Senge, P.M. (1992). The Fifth Discipline, the art and practice of the learning organization. London, Century Business.

Shaw, W.H. And Barry, V. (1989). Moral issues in business. 4th edition. Belmont, California, Wadsworth Pub. Co.

Simon, H.A. (1960). The new science of management decision. Harper & Row, New York.

Van der Spek, R. and Spijkervet, A. (1997) Knowledge Management: Dealing Intelligently with Knowledge. Pub: Knowledge Management Network, Kennisecentrum CIBIT, The Netherlands.

Von Bertalanffy, L. (1950). The theory of open systems in physics and biology. Science, Vol. 111, January 1950, pp. 23-29.

Von Bertalanffy, L. (1968). General systems theory. George Braziller, New York.

Wickens, C. (1992). Engineering psychology and human performance. 2nd edition. New York, NY, Harper Collins Publishers.

Zajac, E.J. (1998). Comentary on ' Alliances and Networks".. Strategic management Journal, Vol. 19, pp. 318-321.

Printed and published by the Health and Safety Executive
C1 03/01

HSE BOOKS

MAIL ORDER
HSE priced and free
publications are
available from:
HSE Books
PO Box 1999
Sudbury
Suffolk CO10 2WA
Tel: 01787 881165
Fax: 01787 313995
Website: www.hsebooks.co.uk

RETAIL
HSE priced publications
are available from
good booksellers

HEALTH AND SAFETY ENQUIRIES
HSE InfoLine
Tel: 08701 545500
or write to:
HSE Information Centre
Broad Lane
Sheffield S3 7HQ
Website: www.hse.gov.uk

CRR 330

£10.00

ISBN 0-7176-1990-7

9 780717 619900