Systemic Failure Modes:
A Model for Perrow's Normal Accidents in Complex, Safety Critical Systems

R.J. Collins[1] and R.Thompson[2]

[1]Sentient Systems Limited, 1 Church Street, The Square, Wimborne, Dorset, BH21 1JH, UK
[2]17 Belmore Road, Lymington, Hants, SO41 3NU

ABSTRACT

In 1984 Charles Perrow produced the book 'Normal Accidents' which presented the 'Normal Accident Hypothesis' (NAH). The NAH states that there are inevitable failure modes of complex, highly coupled systems that are not predictable and hence not preventable. Complexity Theory provides us with tools for understanding systems that exhibit *emergent behaviour*; in other words, behaviours that arise spontaneously as a product of complexity and do not admit to reductionist explanations. The authors argue that complexity theory provides a framework within which the NAH can be understood. The authors introduce the term 'Systemic Failure Modes' (SFM) which are defined to be undesirable behaviours of systems that emerge as a function of system complexity and which are not reducible to smaller (more atomic) constituent components. SFM form a sub-set of Perrow's Normal Accidents and are accessible to the production of testable models. The authors review the fields of complexity theory and of general systems theory and re-state Perrow's NAH in these terms.

KEYWORDS Normal Accidents, Complex Systems, Safety, Systemic, Chaos

1. INTRODUCTION: THE NORMAL ACCIDENT HYPOTHESIS

In 1984 Charles Perrow produced the book *Normal Accidents: Living with High-Risk Technologies*. The book has been most commonly referenced as a large collection of case-studies of accidents, (for example see Reason (1990) and Tenner (1996)). The book does contain a wide range of case-study material relating to accidents and disasters, however, we have been able to find no reference that treats the central thesis of the text. It is Perrow's central thesis, the 'Normal Accident Hypothesis', with which this paper deals.

The Normal Accident Hypothesis (NAH) is succinctly expressed in Perrow's own words:

> "[There are] characteristics of high-risk technologies that suggest that no matter how effective conventional safety devices are, there is a form of accidents that is inevitable"

and...

> "If interactive complexity and tight coupling - system characteristics - inevitably will produce an accident, I believe we are justified in calling it a 'normal accident', or a 'system

> accident'. The odd term 'normal accident' is meant to signal that, given the system characteristics, multiple and unexpected interactions of failures are inevitable. This is an expression of an integral characteristic of the system, not a statement of frequency."

The Normal Accident Hypothesis is controversial indeed and this may explain the failure of other authors to engage with this subject. We consider that it does merit analysis since the conclusion that Perrow draws from it is of such great importance. We believe that efforts should be made to either debunk the hypothesis and move on, or otherwise to consider the implications that would arise from the hypothesis being substantiated.

There are other features of the book "Normal Accidents" that may have prevented it from being more fully addressed by the scientific and engineering communities. Firstly, an attempt is made by Perrow to substantiate the Normal Accident Hypothesis by reference to a large body of case study material. The readings are certainly suggestive that some untreated and unpreventable mechanism underlies a wide range of accidents. However suggestion is not proof. Perrow is a Professor of Sociology and it has been pointed out that a major difference between the social sciences and the physical sciences is the differing emphasis on empirical and observational evidence (Shipman 1982). Problems in social sciences do not admit easily to experimentation and thus sociologist may be constrained to present many arguments on the basis of observation and inferential explanation. Such explanations are less familiar and less convincing to scientists and engineers more used to empirical evidence.

In this paper we will attempt to provide some analytical substantiation for the Normal Accident Hypothesis. Whereas Perrow has provided the high-level, indicators of the phenomenon, we intend to indicate some low-level mechanisms by which it might arise. We shall attempt to provide supporting evidence for the hypothesis based on the emerging field of complexity theory.

Our assertion is somewhat stronger even than Perrow's. His view seems to be that in complex, highly coupled systems, the inevitable faults of component parts are bound to precipitate in unpredictable, and ultimately catastrophic ways. We include these cases but consider also the situations in which no individual component failure occurs and yet the system as a whole manifests a 'pathological' failure behaviour.

In the following section we consider the history of thought concerning systems and the emerging field of complexity theory. We review the arguments supporting the notion that there may well be emergent properties of systems (in this case failure mechanism) that are not traceable to the individual system components.

2. SYSTEM THINKING

The well-worn phrase 'the whole is more than the sum of its parts' occurs frequently in texts on systems theory and design (von Bertalanffy 1968; Meister 1991) and has been attributed by Koestler (1978) to have originated from Smuts:

> "A whole, which is *more* than the sum of its parts, has something internal, some inwardness of structure and function, some specific inner relations, some internality of character of nature which constitutes that *more'* (Smuts 1926)

Mattessich (1982) has argued that the 'Systems' approach as a direct embodiment of the holistic paradigms of philosophers as diverse as Lao-tse, Heraclitus, Leibniz, Vico, Hegel, Marx, Whitehead, Driesch and others. We shall attempt to deconstruct the somewhat flowery language of Smutts and attempt to re-interpret it in line with an analysis of complex systems. We shall explore the nature of the 'more' that Smuts refers to in the context of safety and failure analysis. Smuts attempted to express this concept algebraically, and we would reformulate his notion thus:

$$Whole = \sum_{i=1}^{n} parts_i + x \qquad (1)$$

A key argument to our thesis is that the 'x' in Eqn. 1 does exist in some meaningful, demonstrateable manner.

The general notion that systems do exhibit properties over and above those of their constituent parts has been widely attributed to Hegel[1]. This notion includes the view that the parts of a system cannot be understood in isolation to the whole.

3. REDUCTIONISM AND DECOMPOSITION

A distinction must be made between 'decomposition' and 'reduction'. To decompose is to separate or resolve into constituent parts or elements. Reduction is the process of describing a phenomenon in terms of more 'basic' or 'primitive' phenomena *to which the first is then considered to be equivalent*. Analysis, as we see it, is often decompositional and yet not always successfully reductive. We will argue that something important may be lost in the process of decomposition, such that the products are not a true reduction of the whole.

When we use the word 'analysis' in the sense of 'safety analysis' it is likely that we mean to connote the general process of determining errors and the possibility and consequences of faults. However, 'analysis' has been taken to have an operational meaning:

> "'Analytical procedure' means that an entity investigated be resolved into, and hence can be constituted or reconstituted from, the parts put together" (von Bertalanffy 1968)

That is, 'analysis' is by definition reductionist. But as Bertalanffy goes on to say:

> "Application of the analytical procedure depends on two conditions. The first is that interactions between 'parts' be non-existent or weak...The second condition is that the relations describing the behaviour of the parts be linear...These conditions are not fulfilled in the entities called systems"

If this view is true then the term "System Analysis" is at best an oxymoron and potentially an example of an atomically reduced self-contradictory statement.

4. REDUCTIONISM AND DECOMPOSITION IN ENGINEERING ANALYSIS

System safety analysis invariably follows a reductionist, decompositional approach. Fault-trees are an exemplar of this general rule. Beginning with specific potential behaviours of a system they are used to reason root causes through a process of decomposition. However, only in certain, well defined cases does this decomposition encompass the combination or interaction of failure modes. An exhaustive search of potential failure cases would resemble a maximal fault-tree in which all combinations of all possible failure modes were considered. Clearly this is not practical for any but the most trivial system. The argument for this is that the 'combinatorial explosion' problem resists analysis of permutations of failures.

It has been argued that both Fault Tree Analysis (FTA) and Failure Modes and Effects Criticality Analysis (FMECA) exhibit both inductive and deductive properties (Collins and Leathley 1995). The reasoning processes involved in these analyses involve generalisation and specialisation steps to varying degrees. More importantly the success of both types of analysis are predicated on the applicability of a reductionist approach.

The other major techniques of safety analysis are essentially reductionist: Zonal Analysis (ZA) considers the physical proximity of components and the unintended energy or information flow between them. Common

[1] We have been unable to find an accessible reference to substantiate this attribution. It would appear (rather self-referentially) that the concept of "whole greater than parts" is not expressed in this form in the writings of Hegel. Rather it seems to be encoded in some complex, global manner not accessible to the reductionist techniques available to non-philosophers.

mode failure analysis (CMFA) considers shared susceptibilities between components. Systematic design faults are considered, but as mentioned previously, the use of the word ‘systematic’ refers to the process of design and the ubiquitous occurrence of the error or fault in diverse parts of a system rather than a ‘pathology’ of the system structure as a whole. Such analyses can intersect the functional or failure domain hierarchies considered by Fault Tree Analysis (FTA) and FMECA but they still act in a reductionist manner since they impose their own structural hierarchies. For ZA this hierarchy is one of physical proximity. For CMFA the hierarchies may be of location, of design, of manufacture etc.

It should also be considered that not only are these techniques characteristically reductionist they also operate within the domain of failure itself. We can at least conceive of a system failure that is a product of the interaction of the component parts of a complex system under conditions such that none of the individual components have failed. It is this class of high-level failures that we term ‘Systemic’.

It is not surprising that reductionist, decompositional thinking predominates in system safety analysis since they have dominated science as a whole since the 17th century (Meister 1991). However, there is a rejuvenation of interest in more holistic concepts of systems as a result of pressure from problems that simply do not admit to a reductionist approach (Casti 1979).

We should now consider if any weight can be given to the notion of Systemic failure modes, that is, failures arising at the system level that are not attributable to behaviour (failures) at a lower level of abstraction. If such failures are inconceivable then it will not concern us too much if our analysis methods are exclusively reductive. However, if credence can be given to the concept, then some effort will be required to review the safety analyst’s dependence on decompositional, reductionist techniques.

5. COMPLEXITY AND CHAOS

Casti has referred to the three ‘C’ words of system theory: ‘Connectivity’, ‘Complexity’ and ‘Catastrophe’ (Casti 1979). We take Perrow’s ‘Coupling’ to be synonymous with Casti’s ‘Connectivity’ and add to his list ‘Chaos’ (Casti makes a similar addition in his later work (Casti 1994)).

Casti’s book ‘Complexification’ provides a catalogue of arguments against the reductionist approach:

> “[The] reason for trying to create a science of the complex is to get a handle on the limits of reductionism as a universal problem solving approach” (Casti 1994, pp 273)

Casti describes ‘Complexification’ as the ‘Science of Surprise’ and this relates to our consideration of the Normal Accident Hypothesis. Unpredicted failure modes of systems are surprises indeed.

Complexity theory concerns a number of properties of systems that do not admit to a reductionist approach such as emergent behaviour, chaotic behaviour and ‘deterministic randomness’. It is the recognition of these phenomena that lead to the central argument of this paper; that it may be possible to use the ideas of complexification to build computational models that lend weight to the Normal Accident Hypothesis.

The next sections briefly describe what is meant by the terms ‘Complexity’ and ‘Chaos’ which have a particular bearing on the sections that follow.

5.1 Complexity

Perrow does not provide a formal definition of ‘complexity’ within the book Normal Accidents. ‘Complexity’ has become a much over-used term in recent years and often eludes definition (von Neumann 1966). For example, readers of “Dealing with Complexity”, a text on systems science by Flood and Carson (1988), are left to infer what ‘complexity’ actually means from a series of observations such as “complex situations are often partly or wholly unobservable”.

Within the field of complexity theory the most commonly quoted definition of complexity is that of algorithmic complexity due to Kolmogorov (1965) and Chaitin (1966; 1970; 1974; 1982)[2]. This definition relates specifically to the complexity of a string of binary digits although it can be generalised to other systems. This definition holds that a measure of complexity is provided by the shortest algorithm that can produce the string. In other words it is equivalent to the most dense coding of the information used to describe a particular system.

This definition is attractive at a theoretical level since it leads directly to a number of interesting conclusions in number theory (for example that most real numbers are 'random', i.e. not producable algorithmically by any program significantly shorter than the number itself). However, it is of little practical value to the engineer since this 'measure' of complexity cannot, in general, be computed.

Bennett (1990) has gathered a diverse collection of definitions of the term 'complexity' applicable to physical and biological systems. These definitions refer to such properties as high free energy; the ability of a system to be programmed to perform like a Universal Turing Machine; the existence of long-range order in the system and 'Thermodynamic Depth', the amount of entropy produced in the systems evolution. Each of these is theoretically sound but of little practical value to the engineer who needs a usable metric for complexity.

In the field of Software Engineering the McCabe Complexity Measure (amongst others) is used to measure the complexity of computer programmes (McCabe 1976). This measure is based on a graph representing the control flow of the program. The measure is based on the number of decision paths and loops and is related to the difficulty of testing the program effectively. It seems likely that systems engineers require similar measures of complexity for real physical systems.

In safety related systems an important aspect of complexity is likely to be the 'cognitive' complexity of a system or a situation. This seems only tenuously linked with the pure, mathematical and physical definitions of complexity previously mentioned. For example, a human being may battle for an extended period with a Chinese string puzzle in an attempt to separate one closed loop from another. Topologically however the problem is trivial since the two closed loops are never 'joined' in a mathematical sense. The things that a human being finds hard or is confused by is likely to have as much to do with the function of the human brain as with the domain of the problem.

We can do no more than to alert the reader to the difficulties associated with the term 'complexity'. Like Perrow, we shall side-step this issue and adopt the intuitive, 'dictionary' definition of complexity rather than a formal one.

5.2 Chaos

'Chaos' has been adopted as the short-hand term for the behaviour of non-linear dynamical systems more formally referred to as deterministic unpredictability (Gleick 1987; Hilborn 1994) . Chaos refers to the situation in which the future behaviour of a system is difficult to predict over a long period because it depends on arbitrarily small variations in the current state. Since it is impossible to observe the current state of systems with infinite accuracy, the future behaviours cannot be predicted with accuracy beyond a certain point.

In technical terms, a chaotic system is one in which trajectories through the system state-space diverge exponentially from each other (up to some overall limiting boundary conditions for the system). Such divergence may be measured using the Lyapunov Exponent, a statistic developed to measure chaotic

[2] Curiously, by Chaitin's own admission, this definition was developed and published by Kolmogorov several years before Chaitin published his own work. In a curious reversal of normal academic procedure Chaitin appears to have been adopted as the 'inventor' of this important definition. There seem to be no good reason for this other than that Chaitin is (1) American and (2) Alive, whilst Kolmogorov is (1) Russian and (2) Dead.

behaviour. If two paths that start close together with a separation d_o at time t=0 and the two paths diverge so that their separation at time t satisfies the expression:

$$d(t) = d_o e^{\lambda t} \tag{2}$$

then the parameter λ is called the Lyapunov exponent for the trajectories. If λ is negative then the behaviour is considered to be chaotic.

6. THE IMPLICATIONS OF CHAOTIC BEHAVIOUR

A direct consequence of chaotic behaviour is that the long term future states of systems are impossible to assess from observations of starting conditions and past paths. Although predictions of future behaviour can be made for a short time ahead, the accuracy of predictions reduces rapidly with the period of time for which the prediction is made. Such behaviour might have serious consequences for human operators involved in control of such systems. As systems move into a chaotic region of their behaviour the computational effort associated with 'correct' control decisions increases exponentially. In other words, systems become essentially uncontrollable by the normal mechanisms.

In Normal Accidents Perrow provides a number of case studies of shipping accidents in which the paths of the ships involved were both 'pathological' and clearly unpredictable to the captains involved. These types of case study provide suggestive evidence for systemic failure modes that might be tested through observation and through the production of computational models.

7. CONCLUSION

This paper has reviewed the emerging field of complexity theory and the field general systems theory with respect to system level failure modes. We have termed these failure modes "Systemic" and have argued that they provide a testable explanation of some of the events termed "Normal Accidents" by Perrow.

Central to our thesis has been the consideration of the system level thinking that is characteristic of Perrow's original argument. This has been contrasted to the decompositional, reductionist mode commonly adopted in failure and safety analysis.

The term 'Systemic Failure Mode' refers to system level failure modes that are not products of the failure modes of the constituent components of a system. We have shown that a feature of complex systems is that they may exhibit 'emergent' behaviours. We have argued that pathological behaviour at the system level might also be a potential emergent property of safety critical systems. This is our expression of Perrow's Normal Accidents.

By equating certain system level failures with pathological emergent behaviours of complex systems we have arrived at a position where a testable model might be constructed. If Systematic Failure Modes exist then it should be possible to build computational model that exhibit these types of pathological emergent behaviours. This paper has presented the philosophical and theoretical basis for the existence of Systemic Failure Modes. A second paper by the authors presents a computational models of such failure modes, based on the arguments presented here (Collins and Thompson, 1997).

REFERENCES

Bennett, C.H. (1990). How to Define Complexity in Physics and Why. *Complexity, Entropy and the Physics of Information. SFI Studies in the Sciences of Complexity Vol VIII.* W.H. Zurek (Ed). Addison Wesley: Redwood City, CA.

Casti, J. (1979). *Connectivity, Complexity and Catastrophe in Large-Scale Systems*
John Wiley and Sons : Chichester

Casti, J.L. (1994). *Complexification: Explaining a Paradoxical world through the science of surprise*. Abacus : London

Chaitin, G. (1966). On the Length of Programs for Computing Finite Binary Sequences. *Journal of the Association of Computing Machinery.* **13:4**, 547-569

Chaitin, G. (1970). On the Difficulty of Computations. *IEEE Transactions on Information Theory* **IT-16**, 5-9

Chaitin, G. (1982). Algorithmic Information Theory. *Encyclopaedia of Statistical Sciences*. **Volume 1**, 38-41. Wiley: New York.

Chaitin, G. (1974). Information Theoretic Computational Complexity. *IEEE Transactions on Information Theory* **IT-20**, 10-15

Collins, R.J. and Leathley, B. (1995). Psychological Predispositions to Errors in Safety, Reliability and Failure Analysis. *Safety and Reliability,* **14:3**, 6-42

Collins, R.J. and Thompson, R. (1997). Searching for Systemic Failure Modes. *ESREL 97*

Flood, R.L. and Carson, E.R. (1988). *Dealing with Complexity: An Introduction to the Theory and Application of Systems Science*. Plenum Press: New York.

Garfinkel, A. (1993). Reductionism. In R.Boyd, P.Gasper, and J.D.Trout (Eds) *The Philosophy of science* Massachusetts Institute of Technology : Massachusetts

Gleick, J. (1987). *Chaos: Making a New Science.* Abacus : London

Hilborn, R.C. (1994). *Chaos and non-linear dynamics: An introduction for scientists and engineers*, Oxford University Press: Oxford

Koestler, A. (1978). *Janus. A Summing Up.* Hutchinson and Co. : London

Kolmogorov, A. (1965). *Three approaches to the Definition of the Concept 'amount of information'*. Problemy Peradachi Informatsii

Mattessich, R. (1982). The Systems Approach: Its Variety of Aspects. *Journal of the American Society for Information Science*, **November 1982**, 383-394

McCabe, T. (1976). A Software Complexity Measure. *IEEE Transactions on Software Engineering.* Volume 2. 308-320

Meister, D. (1991). *Psychology of System Design*. Elsevier : Amsterdam

Perrow, C. (1984). *Normal Accidents: Living with High-Risk Technologies.* Basic Books : USA

Reason, J. (1995). *Human Error.* Cambridge University Press : Cambridge, UK.

Shipman, M. (1982). *The Limitations of Social Research* Longman : London

Smuts, (1926). *Holism and Evolution* MacMillan and Co. : London

Tenner, E. (1996). *Why things bite back: New Technology and the Revenge Effect* Fourth Estate, London.

von Bertalanffy, L. (1968). *General Systems Theory: Foundations, development, applications*. Allen Lane, The Penguin Press

von Neumann, J. (1966). *Theory of Self-Reproducing Automata*. University of Illinois Press: Illinois