The Hierarchical Evolution of Socio-Technical Systems

Or

Human Evolution in the 21st Century

Or

Holon a Minute…

By

Todd Ruthman
todd@ruthman.com

A brief history…

For two years (1993-1994), I worked on a Master's Degree Applied Science in Systems Design Engineering at the University of Waterloo doing some systems thinking about systems, the evolution of technology, hierarchies and miscellaneous topics of interest.

This document is the result that never made it to the final review process. I have made several attempts to restart it, and even considered returning to complete my Master's. After 10 years, I finally just decided to convert it warts and all to PDF and post it here.

If you read it and have a moment, please let me know why and what you thought. Thanks.

Introduction

This thesis is about understanding change in human society. The motivation for this thesis came from a research scholarship, provided jointly by the National Science and Engineering Research Council and the Social Sciences and Humanities Research Council, to investigate the concept of sustainable development from a systems design engineering perspective. Development implies progressive change. Before one can hope to achieve sustainable development, then, one must first understand change.

Understanding change is no easy task and there are many approaches one could take. This thesis builds a model of change using the three concepts of system, evolution and hierarchy. Each of these concepts is rich in content and open to many interpretations. The first task of this thesis, then, is to define each of the three concepts so that the reader and the author interpret the terms in the same way. That is the role of its first three chapters. Chapter One looks at systems in general and human systems in particular. Chapter Two examines evolution, and Chapter Three discusses hierarchies. As the chapters progress, they integrate previous material with the new concepts so that by the end of Chapter Three the reader will be familiar with modelling framework of this thesis: the evolving hierarchical system.

Chapter Four applies this modelling framework to explain changes in an arctic, reindeer-herding community in Northern Finland resulting from the introduction of the snowmobile in the early 1960s. From this *specific* example, the chapter extracts *general* processes of change for any evolving hierarchical system. The final chapter, Chapter Five,

closes the thesis by discussing implications this model has for understanding change in human society, as well as possibilities for future work to further the model's development.

The thesis makes extensive use of examples to help clarify the concepts discussed. Hopefully these examples are familiar to most readers. However, readers may find it helpful to relate the material to their own situation, such as their roles in an academic, private or public organization.

The model presented in this thesis is still a work in progress and is more theoretical than applied, more qualitative than quantitative. The goal of this thesis is to lay the foundation of a new perspective on change. It is not the only way to view change, but hopefully it will give the reader a different and useful perspective that complements perspectives the reader may already have. By combining different perspectives of change, Canada may someday achieve its goal of sustainable development.

Chapter One

Concept 1: Systems

1.0 Introduction

This thesis makes several assumptions. The first is that one can model both technology and human socio-technical organizations, such as cities and companies, as systems of interacting components. The term *system* can mean different things to different people. It is important, then, to clarify its definition within the context of this thesis. That is the purpose of this chapter. It consists of three major sections. The first defines the concept of a system in general. The second examines why it is useful, again in a general sense. The last section applies the general system concept to human socio-technical organizations in particular.

1.1 What is a system?

There are many ways to define a system. Jordan, for example, lists 15 different definitions, from which he extracts a core meaning.

> We call a thing a system when we wish to express the fact that the thing is perceived/conceived as consisting of a set of elements, or parts, that are connected to each other by at least one discernable principle.[1]

G.J. Klir represents this definition symbolically as

$$S = (T, R),$$

where "*S*, *T*, *R* denote, respectively, a *system*, a *set of things*, and a *relation* (or, possibly, a set of relations) defined on *T* [Figure 1]."[2] One might instead let *S*, *T*, and *R* respectively denote a

system, things and relationships, since *( )* already represents a set.

Just as there are different ways to define system, there are different ways to define things and relationships. Robert Flood and Ewart Carson provide good definitions in their book, Dealing with Complexity. Things, or elements, "may be anything that is discernable by a noun or noun phrase that all informed observers would agree exists. A **relationship** can be said to exist between A and B if the behavior of either is influenced by the other.... Relationships between elements may be flows of materials, information, or energy."[3]

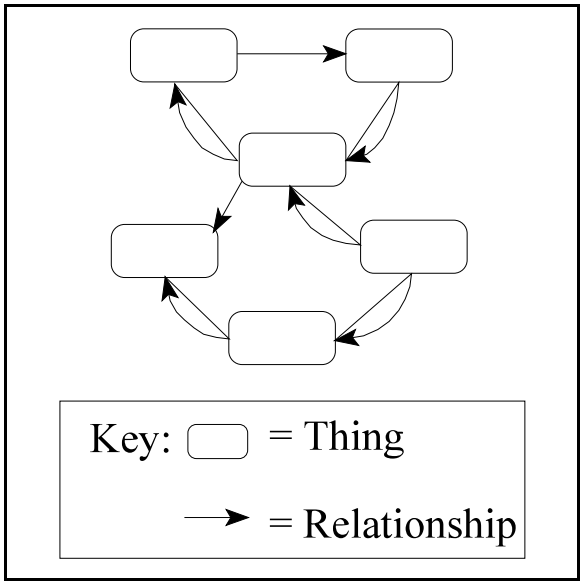

Figure 1: A system

A system so defined can describe many things. People commonly use it to describe large, complex organizations such as the education system and the political system. Some also use it to describe complicated pieces of technology such as a stereo system and the telephone system. In each of these instances, the system is made up of interacting components, of things and their relationships. A system can describe something as small as a human blood cell or as large as a modern city. The definition of system is independent of scale.

While the above are good general definitions of system, they both omit an important conceptual aspect: the system as a *whole*. A system is more than just a collection of things and their relationships. For example, systems architect Eberhardt Rechtin defines a system as "a collection of different things so related as to produce a result greater than what its parts, separately, could produce."[4] Rechtin uses an automobile as an example of a system. An assembled automobile provides mechanised transportation, something none of its parts could do individually. Only by working together *as a whole*, can the parts of an automobile provide mechanised transportation.

In fact, there are at least three aspects to any system description:

1) the system as a whole;

2) the system as an aggregation of interacting parts; and,

3) the system as a part of a larger context.[5]

Describing the system as a whole provides insight into *what* the systems does. Describing the system as an aggregation of parts provides insight into *how* the system does what it does. Finally, describing the system as a part of larger context provides insight into *why* the system does what it does. For example, looking at the telephone system as a whole lets one see that it permits long distance communication between people and machines. Delving into the telephone system as an aggregation of interacting parts lets one see how this is accomplished. Finally, placing the system in its larger socio-economic context lets one see that, among other things, phone companies use the telephone system to make money. Each aspect of the description is important to one's overall understanding of the system. The definitions of Klir and Jordan only consider the second of these three aspects. A more complete definition of a

system, then, may be a whole made up of interacting parts operating in a larger context.

Despite its deficiency, Klir's definition is still useful as a discussion point. Klir observes that the simplicity of the symbolic definition is both its strength and its weakness. "The definition is weak because it is too general and, consequently, of little pragmatic value. It is strong because it encompasses all other, more specific definitions of systems."[6] Its generality lets one distinguish between systems and non-systems, but for practical application the definition needs to be more specific.

This is not to say that the general definition is without merit. It allows one to distinguish between systems and non-systems. This is important because describing something as a system implies that it possesses certain fundamental system properties. Gerald Weinberg, a noted systems writer, defined two of these properties. These are,

a) the Composition Law, and,

b) the Decomposition Law.

a) The Composition Law

The Composition Law states that "the whole is greater than the sum of its parts."[7] Rechtin claims that if the whole is greater than the sum of its parts, then the added value must come from the interactions *between* the parts and not the parts themselves.[8] For example, having a telephone in every household -- the sum of the parts -- is useless unless the telephones can interact. It is the interaction between the parts that gives systems their strength. Metcalfe's Law, named after the inventor of the Ethernet network communications standard, states that the utility of a communication system to a population is "roughly proportional to the number of

users squared."[9] In other words, as the number of people connected to a telephone system or a computer network increases *linearly*, its usefulness increases *exponentially*.

In a true system, the parts are not merely interacting with, but are in fact *dependent* upon, one another. To continue with the telephone example, phone users depend on having access to a complex switching network to call other people. That network, as a private enterprise, in turn depends on having a sufficient number of customers to maintain profitability. In other examples, subways depend on having electricity and electrical utilities depend on customers using electricity. People who travel by car depend on there being an adequate number of sufficiently spaced gas stations while gas station owners depend on travellers to buy gas. One could expand the basic definition of a system, then, from a collection of *interacting* parts which produce a greater result acting as a whole than they could by acting *separately*, to a collection of *interdependent* parts which produce a greater result acting as a whole than they could by acting *independently*. In both cases, the system still operates in a larger context.

Weinberg recognized this aspect of systems and notes that "a system is a collection of parts, no *one* of which can be changed."[10] While this may seem at first a nonsense statement, Weinberg's point is that changing one component in a system has repercussions for the other components, and the system as a whole. Each component depends on the others to remain the same. If one changes, others may need to as well. To continue with the example of a telephone system, telephone companies have expanded their services to include options such as call-display and call-answer. Some of these features work with existing telephones. However, call-display requires a phone with display capability. To take advantage of this service, many customers must upgrade their telephones. Change necessitates change.

b) The Decomposition Law

The lesser known of Weinberg's two laws, the Decomposition Law, is the inverse of the Composition Law. It states that "the part is more than a fraction of the whole."[11] In a system, the interactions and interdependencies between components place constraints on those components, limiting their behaviour. Each component is likely manifesting only a *fraction* of its full potential when part of any given system, that is to say in another context each part can be more than or different than it is in its current context. If the part is removed from one system and placed in another, it may demonstrate *other* aspects of its full potential. Staying with the telephone example, if one connects a phone jack in a household wall to a phone, that person can call and speak with another other person also connected to the system. As long as that jack is connected to the phone, it is constrained to act as a relayer of *voice* communication only. If that person instead connects the jack to a computer modem, he or she can call up bulletin boards or computer networks. In this case, the jack acts as a relayer of *data* communication. The jack's role in the system depends on the other components with which it interacts. Modern personal computers, for instance, take greater advantage of the telephone jack's full potential by combining voice, data and fax capabilities. In each of these cases, the jack manifests only a fraction of its full potential in each of the systems in which it was a part because it is constrained by the other components of the system. In each case, the jack itself is more than a fraction of the whole.

These system properties act a base from which one can develop more specific definitions of systems which have correspondingly more specific associated properties. Klir

notes that there are two independent, yet complementary approaches, to make his definition of system, $S = (T, R)$, more specific. The first restricts T to certain kinds of things, while the second restricts R to certain kinds of relationships. Klir refers to the first method as type (a) and the second as type (b).

> Restrictions of type (a) are exemplified by the traditional classification of science into disciplines and specializations, each focusing on the study of certain kinds of things (physical, chemical, biological, economic, social, etc.) without committing to any particular kind of relations. Since different kinds of things are based on different types of distinctions, they require the use of different senses or measuring instruments and techniques. Hence, this classification is essentially experimentally based.
>
> Restrictions of type (b) lead to fundamentally different classes of systems, each characterized by special kinds of relations, with no commitment to any particular kind of things on which the relations are defined. Since systems characterized by different types of relations require different theoretical treatment, this classification, which is fundamental to systems science, is predominately theoretically based.[12]

For example, consider the system classification schemes developed by (a) Charles Perrow and (b) Kevin Kelly. In each case, the author either restricts T to certain kinds of things, or R to certain kinds of relationships, or both.

a) Perrow's Accident Prone Systems

Charles Perrow developed his conception of systems to understand why some systems seem to fail so spectacularly and so tragically while others do not.[13] His main focus is how different types of systems react to internal failures. He restricts T to human social and technological systems such as assembly lines and nuclear power plants.

Having restricted T, Perrow classifies these specific systems by restricting R to two types of relationships and then ranking systems along two corresponding scales (Figure 2). The

first scale considers whether the interactions between system components are linear, complex or somewhere in between. The second scale considers whether the connections between components are tightly or loosely coupled. Perrow claims that this classification scheme will indicate how that system will react to internal failures.

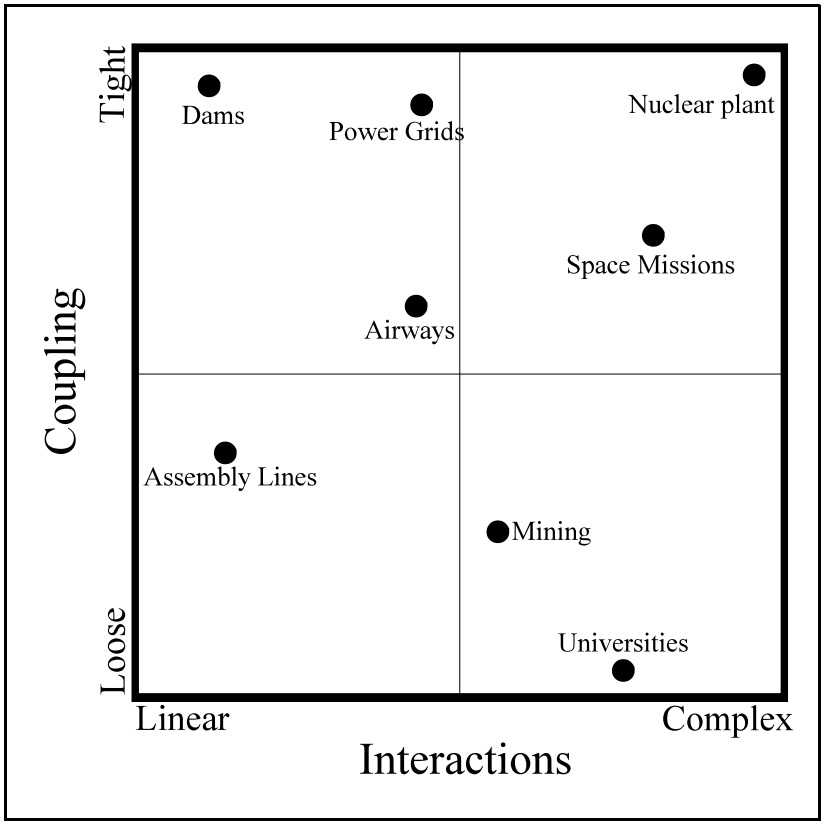

Figure 2: Perrow's Classification of Systems

A system with predominantly *linear* interactions is exemplified by a production line. Each component operates more or less independently in an easily controlled linear sequence. Each component is usually directly connected and influenced by only a few others. When a failure occurs in a linearly connected system, its cause is readily identifiable and its effects are fairly predictable. In contrast, *complex* interactions usually surprise observers. Components are connected to many other components in parallel. The actions of one component therefore

influence many others. Complicated, often hidden, feedback loops emerge with some components affecting others in ways that were never intended. When failure occurs in one component, the cause is rarely obvious and the effects are likely unpredictable because the system is so intertwined. These systems include things like nuclear power plants.

When two components are tightly coupled, what happens to one component directly affects the other. There is no slack, no give, between them. On the other end of the spectrum, components that are loosely coupled have slack, or give, between them. What happens to one does not directly, or at least immediately, affect the other. For example, consider a production line assembling an automobile. The emerging automobile moves through a sequential series of work stations each of which must complete its task before the automobile-to-be can proceed to the next station. If one station were to break down, work at all stations would be suspended until someone repaired the broken station. In this example, the stations on the assembly line are tightly coupled.

A similar, tightly coupled production line may simultaneously be assembling car engines. The end products of these two separate production lines are eventually merged. These two separate lines may be loosely or tightly coupled depending on the manufacturing strategy employed. If a just-in-time strategy is used in which the production of engines must exactly match the production of frames, then the two lines are tightly coupled as a work stoppage in one requires a work stoppage in the other. However, they are not as tightly coupled as the stations *within* each line since one could still produce engines if the engineless automobile production stopped and vice versa. On the other hand, if storage facilities were available to store excess production, then when one line failed the other could maintain

production. In this case, the two lines are loosely coupled. A key property of loosely coupled systems is that if one component fails the effects are not felt throughout the system immediately. In a tightly coupled system, when one component fails the effects quickly ripple through the system.

Perrow claims the ranking of a particular system on one scale is more or less independent of its ranking on the other, so there are four extremes. These are,

a) linear and tightly coupled systems,

b) linear and loosely coupled systems,

c) complex and tightly coupled systems, and,

d) complex and loosely coupled systems.

These four extremes bound all systems. Any system can therefore be represented graphically on a chart generated using these four extremes. See Figure 2, above.

Where a system falls on the chart indicates how that system will respond to internal failures. Of the four extremes, those systems that tend towards the tightly coupled, complex interactions corner pose the most problems for accident management. Since the interactions are complex, only a small subset of the possible failure modes are possible to anticipate. When something does fail, its effects throughout the rest of the system are often unpredictable. Because the components are tightly coupled, the effects of any failure will race through the entire system. People and safety features rarely have time to respond appropriately. Overall, failures in loose, linear systems are much easier to manage successfully.

It does not necessarily follow that tightly coupled, complexly interconnected systems are the most dangerous. This is only the case if the failures are likely to result in actual

suffering, as with nuclear power plants. Perrow's main thesis is that when failures occur, the effects on the system depend greatly on how its components interact. Understanding that helps people to design safer systems. It also helps people understand where they need greater vigilance.

b) Kelly's Clocks and Swarms

Kevin Kelly classifies systems for a very different reason than Perrow. Kelly wants to show that the distinction between purely technological and purely biological systems is becoming more indistinct. Kelly restricts *T* by looking mostly at systems that have large number of interacting components. He restricts *R* by considering only whether component interactions are sequential or parallel. His classification scheme is therefore unidimensional, ranking systems along a scale ranging from totally sequential to totally parallel connections. He calls these extremes clockwork and swarm systems.

> At one extreme, you can construct a system as a long string of sequential operations, such as we do in a meandering factory assembly line. The internal logic of a clock as it measures off time by a complicated parade of movements is the archetype of a sequential system. Most mechanical systems follow the clock.
>
> At the other far extreme, we find many systems ordered as a patchwork of parallel operations, very much as in the neural network of a brain or in a colony of ants. Action in these systems proceeds as a messy cascade of interdependent events. Instead of the discrete ticks of cause and effect that run a clock, a thousand clock springs try to simultaneously run a parallel system. Since there is no chain of command, the particular action of any single spring diffuses into the whole, making it easier for the sum of the whole to overwhelm the parts of the whole. What emerges from the collective is not a series of critical individual actions but a multitude of simultaneous actions whose collective pattern is far more important. This is the swarm model.[14]

A clockwork system is largely sequential and tightly controlled. For example, an

automotive assembly line in which a carefully orchestrated, sequential series of operations produces an automobile to precise specifications is a clockwork system. A swarm system is largely parallel and more constrained by environment than controlled in any way. The obvious example is a beehive, in which the aggregate behaviour of all the individual bees results in a functioning hive.

Some systems are more clockwork in structure while others tend to be more swarm-like. Most human systems are found somewhere in between the two extremes. Kelly contends that technological systems are becoming more and more swarm-like as they become more and more complex. The global economy and telephone system already act in a largely swarm-like manner. Other systems, such as computer software, are following the trend. In contrast, as humans learn more and more about the biological world, they attempt to bend it to their will through genetic engineering and other techniques. The biological world is becoming more clocklike. Kelly suggests that as time goes on, the distinction between clockwork and swarm systems will slowly disappear as all human systems converge to somewhere in between these two extremes.

This trend has both advantages and disadvantages, which Kelly identifies by contrasting the properties of swarm and clockwork systems. Swarm systems tend to be more resilient and adaptive as a whole in the face of change, but this comes at the expense of the individual and the ability to predict and control the system. The strength of a free market economy is its ability to shift the production of goods and services in response to supply and demand. It is flexible. Also, if one manufacturer fails another will often take its place. It is resilient. Both these attributes come at the expense of the individual. There is enough redundancy in the

system that, although the economy as a whole will usually survive, no one person's income is ever guaranteed. Finally, a survey of any financial news shows that predicting and controlling the economy to any significant degree is notoriously difficult, if not impossible.

Clockwork systems, on the other hand, are predictable and therefore more controllable, but they also tend to be inflexible in the face of change. For example, the flow of material through an automotive assembly line is fairly predictable and controllable but it would take a major restructuring for the line to produce anything but automobiles. If the economy suddenly stopped demanding cars and trucks, these factories would have a hard time adjusting, just as the defense industry may have a difficult time adapting to the threat of peace.

Both authors discuss their respective system definitions in much greater detail than presented here, but the above material gives some indication of the different types of system definitions and classifications possible by restricting *T* and *R*. This thesis, for instance, is concerned primarily with hierarchical and evolutionary relationships in human systems. Section 1.3 of this chapter discusses these concepts in more detail.

Are systems real?

The above system definitions are useful for their specific purposes. However, it is legitimate to ask if systems are real, concrete entities or if they are theoretical, abstract constructs. A danger in working with systems is to attribute more objectivity to a system description than it warrants. Whether or not the so-called real world exists independently of human thought is open to philosophical debate. What is certain is that humans have access to

that world only through their observations.

In making an observation, people impose their own values on what they see. What someone sees depends more on *how* and *why* they are looking than on *what* they are looking at. One should not confuse accuracy with objectivity.[15] Several people can *look* at the same system, but *see* it very differently. For example, most home owners likely see the telephone system primarily as a method of communication whereas the telephone company likely sees it as a method to make money. Recall Klir's symbolic general definition of a system as a set of things, *T*, and their relationships, *R*. More specific system definitions involve restrictions on *T* and *R*. Different people may choose different restrictions. Some will focus on thermodynamic relationships between biological organisms, while others may consider predator-prey relationships between the same organisms. This will shape what they see when looking a given ecosystem. Even people choosing the same restrictions may still see different systems depending on their values, beliefs, the tools they use, and the context in which they are looking.

And yet, there is enough similarity between different people's perceptions of these systems to suggest that a system is not an entirely fabricated concept, that it does reflect at least in part what exists in the so-called real world. Jordan comments on this relationship between observer and observed.

> Whenever one person can point to, or explain, a set of elements, and the nature of the connectivity between these elements, to another person, then the other person will perceive/conceive of the set as an entity, a thing. The word 'system' will then spontaneously emerge as the adequate expression, as a proper name, for this thing. A system is therefore an interaction between what is 'out there' and how we organize it 'in here'. 'System' denotes an interaction between the objective world and how it is looked at or thought about; it denotes a mode of perceptuo-cognito organization.[16]

But while system descriptions may reflect in part the objective world, they remain to a large degree arbitrary constructs of the observer. This subjective nature of system descriptions manifests itself in several ways. Two of these have been discussed already -- the restrictions observers place on *T* and *R*. A third, the importance of context, was also discussed briefly. It is the choice of system boundaries, though, that is perhaps the most subjective aspect of all system descriptions.

Describing some aspect of the world as a system requires the observer to choose boundaries. Boundaries separate the system under study from the rest of the world, from its *environment*. The concept of environment lets one differentiate between a closed system and an open system. In a closed system, relationships do *not* exist between the things in the system and the things in its environment. In an open system, these relationships *d*o exist. All systems descriptions are theoretical constructs, but closed systems are even more so as they represent an idealized system which one can separate cleanly from its environment. Few systems lend themselves to this description. Open system descriptions explicitly acknowledge the role of the environment and tend to be more useful.

Distinguishing system from environment is rarely straightforward. The observer must determine which things and relationships are important enough to consider part of the system, and which can be relegated to the environment. Since the boundary between the two is usually ambiguous, some may find it useful to use a graduated approach (Figure3). Using this method, one identifies a narrower system of interest (NSOI), a wider system of interest (WSOI), a narrower environment of interest (NEOI), and a wider environment of interest (WEOI). The phrase 'of interest' explicitly acknowledges the role of the observer. The WSOI, NEOI and

WEOI are in effect larger and larger contexts of the NSOI.

This graduated approach allows the observer some flexibility in separating system from environment. For example, if a thing or relationship seems too important to be part of the environment but not important enough to include in the narrower system of interest, then it can be placed in the wider system of interest. Each person will likely have their own preferred methods for setting system boundaries.

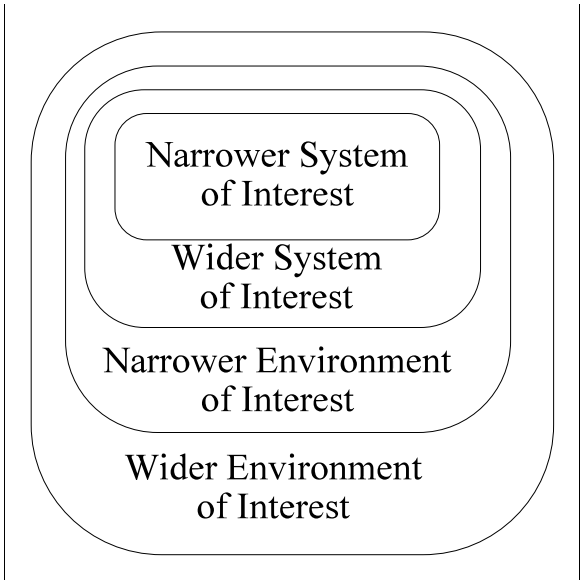

Figure 3: System Boundaries, adapted from Flood and Carson (1968), 8.

During this process, it is often helpful to use system diagrams. System diagrams are "those that show relations as parts to a whole."[17] Just as there are many ways to define systems, there are many different types of systems diagrams. Flood and Carson provide a good overview of system diagrams in their book, Dealing with Complexity.[18] As there are too many to discuss within the confines of this thesis, interested readers are asked to refer to Flood and Carson's book. Because there are so many possible types of system diagrams, Flood and

Carson offer some simple but important rules for their construction.

> **RULE 1.** Define what kind of diagram is appropriate. (Is it to convey structural relationships? Material flows? Information flows? The logical
>
> dependencies among activities? Interaction of process parameters? Social influences?)
>
> **RULE 2.** Decide on a convention for the diagram's entities [things] and relationships, ensuring that a symbol is selected for each kind of element.
>
> **RULE 3.** Provide a key that ensures that different readers will interpret the diagram in the same way.[19]

These rules should help when developing system diagrams. They will be used in later chapters when developing systems diagrams for this thesis.

1.2 Why study systems?

So far this chapter has spent a great deal of time discussing what a system is without explaining why one should be interested. Systems science arose to redress the limitations of traditional science. Warren Weaver, one of the co-founders of information theory, discussed these limits in some detail[20]. He wrote that early science dealt with problems of *simplicity*. However, writing in 1948, he could see that more recent science addressed problems of *disorganized complexity*, and that future of science lay largely in problems of *organized complexity* for which existing science was inappropriate. Problems of simplicity involve a small number of objects interacting in well understood ways which one can analyze using two-variable mathematics, such as Newton's Laws of Motion. Problems of disorganized complexity involve very large numbers of objects interacting in *statistically* understood ways. Finally, problems of organized complexity are found in between these two extremes.

Weaver uses the idealized motion of billiard balls on a pool table to illustrate his distinctions. Modelling the theoretical movement of a single ball on the table is a fairly straightforward application of Newton's Laws of Motion. It involves only a few variables and is a problem of simplicity in Weaver's terminology.

Increasing the number of moving balls, however, changes the nature of the problem. At some point, the problem becomes unmanageable, "not because there is any theoretical difficultly, but just because the actual labour of dealing in specific details with so many variables turns out to be impractical."[21] An interesting thing happens, though, when the number of moving balls gets *sufficiently* large. Beyond a certain threshold, one reaches the realm of disorganized complexity and can usefully apply statistical methods of analysis. Weaver states, though, that for statistical methods to work not only must the number of variables be large but that the individual behaviour of *each* of the many variables must be erratic or possibly even unknown. He notes this as an advantage of statistical methods in that they are *not* "restricted to situations where the scientific theory of the individual events is very well known."[22] For example, insurance companies are financially viable not because they necessarily understand the individual behaviour of *each* customer, but because they understand the statistical behaviour of *all* their customers. Finally, he observes that just as the analysis of problems of simplicity works better with small numbers of variables, the analysis of problems of disorganized complexity works better with increasingly larger numbers of variables.

But what of problems of organized complexity? Weaver argues that while the science of his day could analyze problems of simplicity and problems of disorganized complexity, they could not practically analyze problems of organized complexity.

> One is tempted to oversimplify, and say that scientific methodology went from one extreme to the other -- from two variable problems to an astronomical number -- and left untouched a great middle region. The importance of this middle region, moreover, does not depend primarily on the fact that the number of variables is moderate -- large compared to two, but small compared to the number of atoms in a pinch of salt. The problems in this middle region, in fact, will often involve a considerable number of variables. The really important characteristic of the problems of this middle region, which science has as yet little explored or conquered, lies in the fact that these problems, as contrasted with the disorganized situations with which statistics can cope, show the essential feature of *organization*.[23]

It is the large number of variables that makes analysis with tools such as differential equations impractical. It is their organization that makes analysis with statistical methods impossible.

Weaver hoped for a third great advance in science, the first two being two-variable analysis and statistical analysis, which could address problems of organized complexity. He saw promise in two war-time advancements: computers and operations analysis. Weaver prophesied that computers "will make it possible to deal with problems which previously were too complicated, and more importantly, they will justify and inspire the development of new methods of analysis applicable to these new problems of organized complexity."[24] In the second approach, namely operations analysis, multi-disciplinary teams pooled their knowledge and perspectives to analyze problems of organized complexity, such as transporting troops and supplies across the Atlantic during a war, and came up with the best solution they could.

With the ability to do rapid, complicated calculations and store vast amounts of information, computers have indeed made inroads into the realm of organized complexity in the last 50 years. The overall result, however, has simply been to re-adjust the boundaries separating the three classes of problems, increasing the number of problems classified as simple or as disorganized complexity. The challenge of organized complexity remains. Multi-

disciplinary teams, on the other hand, are still one of the best ways to address problems of organized complexity.

More recently, Gerald Weinberg represented Weaver's classifications graphically (Figure 4). He describes problems of disorganized complexity as the realm of aggregates, problems of simplicity as the realm of machines and things that are machine-like such as planetary orbits, and problems of organized complexity as the realm of *systems*. Weinberg contends that systems thinking holds the answers for problems of organized complexity. This thesis, for example uses systems theory to analyze the evolution of human technology and organizations. Ironically, or perhaps intentionally on the part of Weaver, systems science emerged roughly at the same time that his article called for a third great development in science.

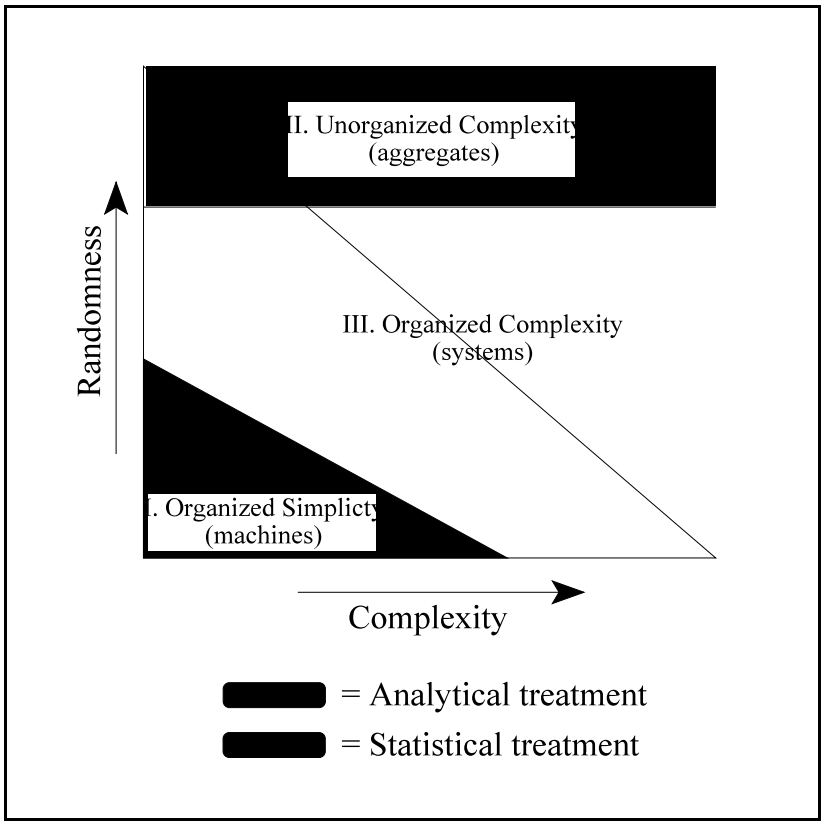

Figure 4: Weinberg's Graphical Representation, from Weinberg (1975), 18.

Systems thinking has a long history. Some specific concepts now associated with systems thinking, such as what we currently know as cybernetics, have actually been around for centuries. In fact, Ludwig von Bertalanffy attributes the systems maxim "the whole is more than the sum of its parts" to Aristotle.[25] It is Bertalanffy, however, whom people recognize as the father of systems thinking, as the person who first consolidated its diverse ideas.[26] Bertalanffy first formulated the notion of his General Systems Theory orally in the 1930's and in publications after World War II, but it actually germinated from his earlier work in organismic biology in the late twenties.[27] According to Flood and Carson, "Bertalanffy envisaged a framework of concepts and theory that would be equally applicable to many fields of interest."[28]

General Systems Theory is "based on the idea that homologies exist between disciplines that have traditionally been considered as being separated by their different subject matters."[29] To understand this, Bertalanffy suggests that we

> take a simple example, an exponential law of growth applies to certain bacterial cells, to populations of bacteria, of animals or humans, and to the progress of scientific research measured by the number of publications in genetics or science in general. The entities in question, such as bacteria, animals, men, books, etc., are completely different, and so are the causal mechanisms involved. Nevertheless, the mathematical law is the same. Or there are systems of equations describing the competition of animal and plant species in nature. But it appears that the same systems of equations apply in certain fields in physical chemistry and economics as well. This correspondence is due to the fact that the entities concerned can be considered, in certain respects, as "systems," i.e. complexes of elements standing in interaction.[30]

He notes that "there appear to exist general system laws which apply to any system of a certain type, irrespective of the particular properties of the system and of the elements involved."[31]

Bertalanffy was not alone in his thoughts. Kenneth Boulding reached similar

conclusions through his work in economics and social science, originally referring to these ideas as general empirical theory. Bertalanffy, Boulding and other like-minded individuals founded the Society for General System Research in 1954.[32] When formed, the major functions of the society were to:

> (1) investigate the isomorphy of concepts, laws, and models in various fields, and to help in useful transfers from one field to another; (2) encourage the development of adequate theoretical models in the fields which lack them; (3) minimize the duplication of theoretical effort in different fields; (4) promote the unity of science through improving communication among specialists.[33]

General Systems Theory emerged from research in basic science in general, and biology in particular. In fact, all the founders of the Society for General System Research, except for Boulding, were involved in biological research.[34] In contrast, the modern fields of cybernetics, operations research and management science resulted from the much more *applied* logistic and resource management challenges of World War II. So while World War II may have "acted as a catalyst for the growth of systems science,"[35] Bertalanffy asserts that *his* version of modern systems theory emerged independently from that war effort but he does concede that cybernetics and related approaches "showed many parallelisms with general systems theory."[36]

Systems thinking originated from, and has since expanding into, many diverse disciplines. Figure 5 models the general development of systems thinking, showing the different processes involved in its evolution. The work in this thesis lies primarily in the first development cycle in which systems thinking, when formalized, leads to systems theory which promotes, in turn, systems thinking. However, as will be seen in subsequent chapters, it draws from other disciplines, most notably evolutionary theory and hierarchy theory.

Systems thinking helps one analyze all three types of problems, namely those of

simplicity, of disorganized complexity and of organized complexity. However, it is particularly useful in helping to understand problems of organized complexity. Since these system have too many components to analyze with calculus, one must cease striving for strict predictability of component behaviour. Since these systems have too few components or are too organized to analyze statistically, one must give up strict predictability of the whole. The identification of system properties will, however, provide some qualitative understanding of the system behaviour. This thesis develops some general system properties to help in the analysis of change in human socio-technical systems.

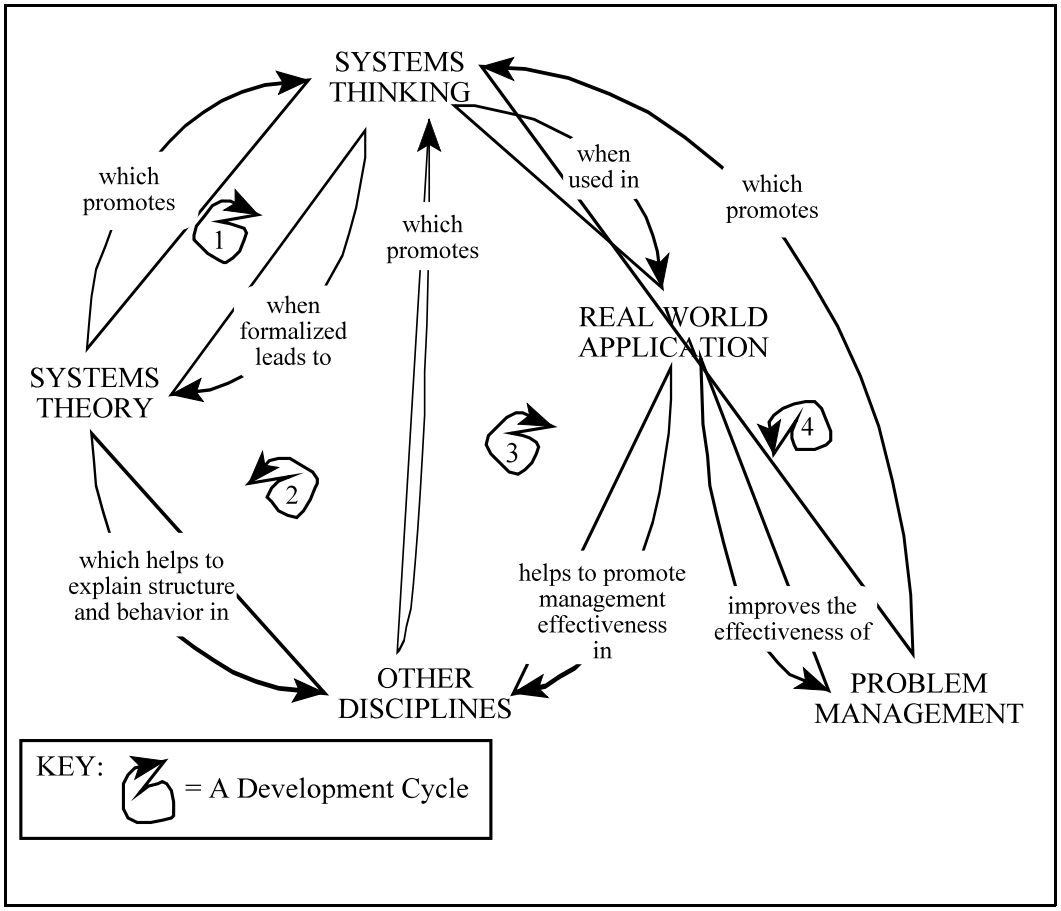

Figure 5: Development of Systems Thinking, from Flood and Carson (1988), 3.

1.3 Human Socio-Technical Systems

A key idea in this thesis is the concept of a human socio-technical system. As it stands, this is a rather vague term and requires clarification. To review, a system is a whole made up of interacting parts and existing within a larger context. In human socio-technical systems, these parts are humans and their tools. This thesis, then, restricts *T* to humans and their tools.

Tool, like system, is a rather all-encompassing term. This thesis defines a tool as anything a person can *use* to achieve an objective. For example, this author uses his computer to write this thesis, his job as a teaching assistant to generate income, the supermarket to supply food, and his apartment to provide shelter. The verb *use* transforms its object into a tool. A rock becomes a weapon when thrown; a log becomes a bridge when lain across a chasm. This thesis, then, restricts *R* to use. It examines how humans use their tools. Please note that these tools do not need to by physical.

Tools may be classified into different categories, though any such attempt likely has some exceptions. For example, one could place tools into the following seven broad categories:

a) physical artifacts and the human body;

b) mental artifacts and the human mind;

c) abstract artifacts;

d) other humans;

e) organizations;

f) non-human biological matter and energy; and,

g) abiological matter and energy.

Each of these categories needs explanation.

a) Physical Artifacts and the Human Body

Physical artifacts are those objects most people would readily refer to as tools.[37] These are human-made concrete objects such as hammers, computers and robots. Humans can also use their body, in part or in whole, as a tool. For example, people use their hands to grasp objects and protesters sometimes use their bodies as barriers. This category includes the human body because it is a physical object. One can argue that the human body was the first tool made by humans.

b) Mental Artifacts and the Human Mind

Mental artifacts are human-made tools used by the mind.[38] They include such things as mathematics, logic and language. In evolutionary terms, mental artifacts such as mathematics are fairly recent developments. Long before humans developed such formalized tools, however, the human mind was still an important tool. The knowledge, skill and instinct of the human mind complements the human body as a tool for survival.

Physical artifacts often encapsulate mental artifacts, eliminating the necessity for people to learn mental artifacts in order to use them. For example, software packages such as Maple and Mathematica can now solve calculus problems symbolically. Symbolic calculus, once only a mental artifact, is now available as a physical artifact as well. Similarly, instant cake mixes encapsulate knowledge of chemistry and recipes. This encapsulation is an ongoing process.

c) Abstract Artifacts

This category exists to classify concepts like jobs as tools. A job in neither a physical nor a mental artifact, but people use jobs as tools to generate income, among other things. One could argue that people do not actually use jobs to make money, but rather their mental and physical artifacts. This category is included more for completeness than for necessity.

d) Other Humans

People use not only artifacts but each other as tools. For example, when water pipes fail in this author's apartment, he calls a plumber. Rather than do it himself, the author uses that plumber as a tool to fix his broken water pipes. Similarly, he uses his professors as tools to gain knowledge. The author uses the plumber as a tool because that person is skilled in plumbing while the author is not. Specialized knowledge in humans is another form of encapsulation. As long as the author has money available, he does not need to learn how to fix his pipes because he can hire someone else to do so for him.

Classifying people as tools may seem dehumanizing since it values them instrumentally, as a means to an end. There are at least two reasons why this is not necessarily the case. The first has to do with exchange of services between people. The second has to do with human nature.

First of all, there is often a reciprocity in using people as tools. The author may use the plumber as tool to fix his pipes but, in exchange, the plumber uses the author as a tool to provide income. This reciprocity can turn what could be an exploitive, parasitic relationship into a more cooperative, symbiotic one. Of course, it does not have to.

Secondly, using someone as a tool and valuing that person instrumentally does not preclude also valuing that person intrinsically as a human being. Even though the author uses the plumber as a tool, he can still consider the plumber as a human being. However, conflict between valuing the plumber instrumentally and intrinsically is possible, especially when negotiating fees.

The concept of humans as tools in not new. For example, William Kuhns notes that Lewis Mumford describes "the immense organizations of slaves that built the pyramids as megamachines: machines that comprised men as their moving parts."[39] Machines are tools, in this case tools to build pyramids. This leads directly to the classification of organizations as tools.

e) Organizations

This thesis defines organizations as systems of people and their physical and mental artifacts. If people can use humans, physical artifacts and mental artifacts as tools, then they can certainly use organizations of all three as tools. This author, for example, used the Social Sciences and Humanities Research Council of Canada as a tool to generate funding for this thesis. He is currently using the Department of Systems Design Engineering to help complete his thesis.

f) Non-Human Biological Matter and Energy

God may or may not have given humans dominion over the other creatures on this planet. It is an issue open to moral and theological debate. However, everyday people use

plants, animals and other biological species for food, shelter and many other purposes. When people use them, they become tools. Again, however, this does not preclude valuing nature intrinsically as well.

g) Abiological Matter and Energy

This last category includes all things not biological and not made by humans, such as iron, uranium, and stone.

The above classification scheme is not meant to be taken as absolute. In many ways, it is more for illustrative than practical purposes, to give an indication of all the different types of tools possible. However, it is helpful for that same reason.

By working together as a system, humans and their tools can achieve things they could not acting independently. The whole is greater than the sum of its parts. For example, the spoken language is a powerful tool by itself. But combine language with ink and paper, and one can do things impossible with the spoken word. By working together in larger socio-technical systems, humans can accomplish even greater tasks. As part of a such a system, each human will likely both use others and be used by others as tools. Recognizing these reciprocal relationships is a necessary part of analyzing human socio-technical systems.

In trying to understand these systems, one needs to use the three levels discussed earlier in this chapter. Looking at humans and how they use their tools, the things and relationships of the system, is one level of systems analysis. It helps one understand how the socio-technical system works. To understand what it is doing, one must analyze the system as a whole. For

example, one might consider the Department of Systems Design Engineering as whole and see that it, among other things, transforms high school graduates into Bachelors of Applied Science. But to understand why a system does what it does, one must look at the larger context within which that system exists.

Humans use tools to satisfy their needs within a given environment, or context. These needs change depending on the context. For example, psychologist Abraham Maslow describes a hierarchy of needs in which "needs at the lower level of the hierarchy dominate an individual's motivation as long as they are unsatisfied. Once these are adequately satisfied, the higher needs occupy the individual's attention [Figure 6]."[40]

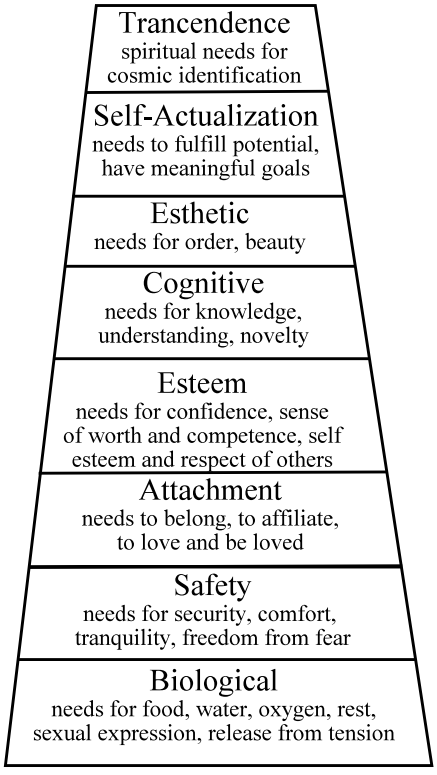

Figure 6: Maslow's Hierarchy of Needs, after Zimbardo (1992), 433.

Whether or not one agrees with Maslow's hierarchy is largely irrelevant. The point is that humans have needs which they try to satisfy and these needs relate to the context in which humans find themselves. In trying to satisfy these needs, they make use of the many different types of tools categorized above.

Each individual will use a different collection of tools. Since there are many different types of tools available, there is effectively an infinite number of possible combinations. This thesis will refer to a particular individual's selection of tools as that person's tool set and use systems theory to analyze them.

1.4 Summary

In summary, there are many different ways to define a system. At its core, a system is a set of things and their relationships. This is only one aspect of a system, however. A system description must include analysis of the system as a whole, as an aggregation of interacting parts, and as part of a larger context. A system, then, is a whole made up of interdependent parts which operate in a larger context. One can make this general definition more specific by restricting the possible set of things and relationships in the system. These restrictions are two ways in which system definitions become more subjective. Context also contributes to their subjectivity. Finally, the necessary process of setting boundaries to separate system from context ensures their subjectivity. Systems theory helps in the analysis of problems of organized complexity which defeat the traditional tools of science. It has a long history but has gained significant recognition only since the mid forties. One can use systems theory to analyze human socio-technical organizations by looking at how humans use tools to achieve

their objectives in a given context. Subsequent chapters expand on this idea by incorporating the concepts of evolution and hierarchy.

1. N. Jordan, "Some Thinking about 'System'," in Systems Thinking Volume Two, ed. F.E. Emery, (Markam, Canada: Penguin, 1981), 24.

2. G. J. Klir, "Systems Science: A Guided Tour," Journal of Biological Systems 1 (1993): 27.

3. Robert L. Flood and Ewart R. Carson, Dealing with Complexity (New York: Plenum Press, 1988), 7.

4. Eberhardt Rechtin, "The art of systems architecting," IEEE Spectrum (October 1992): 66.

5. I.V. Blauberg, V.N. Sadovsky, and E.G. Yudi, Systems Theory (Moscow: Progress, 1977), 132-133.

6. Klir, 28.

7. Gerald M. Weinberg, An Introduction to General Systems Thinking (Toronto: John Wiley & Sons, 1975), 43.

8. Rechtin, 66.

9. Staff writer, "The accidental superhighway," Insert section, The Economist: A Survey of The Internet in The Economist (July 1st 1995): 4.

10. Weinberg, 162.

11. Weinberg, 43.

12. Klir, 31.

13. Charles Perrow, Normal Accidents (New York: Basic Books, 1984), Chapter 3.

14. Kevin Kelly, Out of Control: The Rise of Neo-Biological Civilization (Don Mills, Ontario: Addison-Wesley Publishing Company, 1994), 21.

15. Tim Allen, "The Ecosystem Approach," seminar given at the University of Waterloo, October 2, 1995.

16. Jordan, 24-25.

17. Flood and Carson, 36.

18. Flood and Carson, 35-54.

19. Flood and Carson, 38.

20. Warren Weaver, "Science and Complexity," American Scientist 36 (1948): 536-544.

21. Weaver, 537.

22. Weaver, 538.

23. Weaver, 539.

24. Weaver, 541.

25. Ludwig von Bertalanffy, Perspectives on General System Theory (New York: George Braziller, 1968), 147. Hereafter referred to as Bertalanffy 1968a.

26. Flood and Carson, 2.

27. Bertalanffy 1968a, 152-153.

28. Flood and Carson, 2.

29. Flood and Carson, 2.

30. Ludwig von Bertalanffy, General Systems Theory (New York: George Braziller, 1968), 33. Hereafter referred to as Bertalanffy 1968b.

31. Bertalanffy 1968b, 37.

32. Bertalanffy 1968a, 155.

33. Bertalanffy 1968a, 155.

34. Klir, 53.

35. Flood and Carson, 3.

36. Bertalanffy 1968a, 156.

37. Donald A. Norman, Things That Make Us Smart (Reading, Massachusetts: Addison-Wesley, 1993), 3-17.

38. Ibid.

39. William Kuhns, The Post-Industrial Prophets (New York: Weybright and Talley, 1971), 6.

40. Philip G. Zimbardo, Psychology and Life, 13th ed., (New York: HarperCollins, 1992), 433.

Chapter Two

Concept 2: Evolution

2.0 Introduction

The previous chapter examined the first major assumption of this thesis, namely that one can make system models of both technology and human socio-technical organizations, such as cities and companies. Trace the history of any one particular system, say for example the University of Waterloo, and one will likely find that the system has changed significantly from its original form. The second major assumption of this thesis is that one can model such changes using evolution.

Evolutionary theory emerged in the 1800s to explain the biological changes in species suggested by the fossil record. The core idea in biological evolution is that

> organisms and their environment have an intimate and dynamic relationship: Plants and animals are **adapted** to their environments, that is, they possess physical characteristics and patterns of behavior that help them survive under a given set of natural circumstances. When environments change (as the geological record shows they continually do), organisms must alter their adaptive characteristics if they are to survive.[1]

The specific processes originally used to explain this dynamic relationship between organism and environment have since undergone many changes and refinements and in fact many are still open to debate, but this core idea of evolution remains even today.

Evolutionary theory quickly outgrew its biological roots. Today, people often speak of

social and technological change in evolutionary terms as well. Is this appropriate? Can one take an idea rooted in biology and extend it to the non-biological? Answering these questions requires a closer look at evolution. That is the purpose of this chapter. It has four major sections. The first shows evolution to be a general theory applicable to many different fields. The second discusses why the study of evolution is important. The third briefly looks at evolution in a systems context. The last section then looks at evolution in the context of human socio-technical systems.

2.1 What is Evolution?

Evolution is a theory which explains the sequential change of a given entity's structure and/or behaviour over time as it adapts to its environment. The general theory finds use in many different fields. In biology, the entity could be a species of primate, insect or paramecium. In cultural anthropology, the entity could be a village, a city or a civilization. In civil engineering, the entity could be a style of bridge, the composition of concrete or a method of construction. On the surface, it may seem inappropriate to use a similar theory to explain change in widely different fields such as ecology and economics. Such general theories, however, are at the heart of systems thinking. While it may be true that in each field the *specific mechanisms* of evolution are different, the *general concept* of evolution remains the same. In ecology, it might be genetic mutation and natural selection. In economics, it might be entrepreneurship and market competition. But in both cases, entities change in a sequential manner over time as they adapt to

their environment.

Since people apply evolution to so many different fields, it is important to differentiate between those aspects common to all applications and those aspects specific to individual applications. For instance, evolution often implies *generational* change. That is, it traces the changes from one generation of the entity to the next. Generational change, however, is not required for an entity to evolve. The government of Canada, for example, has existed for over a century but has evolved a great deal since its inception in 1867. Therefore, while change may be necessary for evolution, it need not be generational change. This thesis is interested in identifying those aspects of evolution common to all applications.

There have been many attempts to develop a general theory of evolution applicable to any type of system, be it biological, social, technological or anywhere in between. The journal World Futures reports exclusively on the development of these models. The reader is referred to the works of Ervin Laszlo listed in this thesis' bibliography for a good overview of the more advanced of these general theories.

While these theories are interesting, it is important to step back further to look at the *fundamental* concepts of evolution. These concepts are:

a) evolutionary units;

b) evolutionary context;

c) innovation; and,

d) selection.

Each term needs further clarification.

A) Evolutionary Units

The core idea of evolution, as quoted in the introduction to this chapter, describes the adaptive relationship between an organism and its environment. Both organism and environment are necessary for evolution. The term organism, though, is specific to biology. Its analogue in economics might by a company or a product. In each instance, *something* is evolving. This thesis chooses to call that something an evolutionary unit. Evolution, then, describes the adaptive relationship between an evolutionary unit and its environment.

As it stands, the term is rather vague. This is necessary in part to maintain its generality, but what exactly is an evolutionary unit? Consider for a moment all the different things that one might say evolve. These include biological species such as humans and ants, socio-economic organizations such as companies and cities, or even technological artifacts such as automobiles and televisions. In each instance, the unit in question is an entity made up of interacting parts and existing in a larger context. An evolutionary unit, then, is a system. Since there are very few things that cannot be described as a system, there are few limits upon what can be considered a unit of evolution.

The previous chapter indicated that systems are human constructs, models resulting from the interplay of human observation and objective reality. If an evolutionary unit is a system, the same must hold true for it. Evolutionary units, like systems, are in the eye of the observer. A

person's choice of evolutionary units, then, should be made explicit.

B) Evolutionary Context

The general term evolutionary unit replaces the specific term organism, but what of the term environment? As was seen in the previous chapter, environment has already taken on a meaning more general than its biological one. Environment, though, has a passive connotation that belies the importance of its role in evolution. It is often taken as just there. Also, even though it is commonly used in describing social and technological systems, it has not entirely escaped its ecological origins. For these two reasons, this thesis prefers the term evolutionary context.

Everything evolves within a specific context. Without reference to a specific context, all observations are meaningless. For example, evolution is often associated with the concept of "survival of the fittest".[2] But fitness is a relative term and depends entirely on the context in which it is evaluated. High performance race cars may succeed in the context of a race track but be totally inappropriate in the context of the highways of a fuel-conserving and safety-conscious society.

Evolution results from the adaptive interplay between an evolutionary unit and its evolutionary context. They cannot be separated. In fact, an evolutionary unit is physically *part* of its context. For example, one cannot understand the evolution of the automobile without considering the larger socio-economic system of which it is a part. An evolutionary analysis of

either the automobile or the socio-economic system separately is meaningless. They must be considered together. An evolutionary context is not an abstract entity in which evolutionary units exist. It *is* the evolutionary units or, more accurately, it is the aggregation of all the interacting evolutionary units it contains. If the evolutionary context is part of an even larger context, then it too can be called a system. The Canadian socio-economic system, as a context, is made up of, among other things, automobiles, people, and social institutions, and is itself part of the larger *global* socio-economic system.

This point is important because if one element in a context changes, this means that the context *itself* has also changed. For example, if the automobile becomes more fuel efficient, then overall fuel consumption in its socio-economic context may drop. This will affect the oil industry which may, or may not, lay off hundreds of employees to compensate. This in turn will have other effects. Nothing exists in isolation. Remembering Weinberg's observation from Chapter One that no *one* component in a *system* can change without affecting others, we now realize no one *unit* in a *context* can change without affecting others. When one element in a given context changes, the others elements in that context may have to adapt to compensate.

C) Innovation

The ability to adapt is crucial to evolution. It is the *adaptive* interplay between evolutionary units and evolutionary context that drives evolution. So far, this chapter has discussed the units and context, but not the adaptive interplay. There are two aspects to this

interplay: innovation and selection. This section deals with innovation. The next discusses selection.

If a unit is to evolve, then some process must exist which can alter its form or behaviour. In biological evolution, this is mainly achieved with genetic processes such as mutation and the reproductive mixing of chromosomes. In technological systems, the task often falls to design engineers to modify existing products or design new ones. Whatever the source, some process of innovation must exist for a unit to evolve. Without innovation, existing units cannot adapt to new selection criteria, nor can new species emerge to challenge the old.

This does not imply that a process of innovation is *required* for a given unit to survive in its context. The cockroach has apparently remained largely unchanged for millions of years. It only implies the such processes are necessary if the unit is to evolve.

D) Selection

Why do some evolutionary units continue to exist while others are extinct? The answer lies in the second part of the adaptive interplay between unit and context, namely selection. Selection processes determine which units continue to exist and which do not. To borrow an analogy from writer Kevin Kelly and biologist Lynn Margulis, innovation is the author of evolution while selection is the editor.[3]

In biological systems, for example, natural selection acts to edit the numerous forms authored by genetic processes, keeping some and discarding others. Processes of natural selection include competition between species and individuals for scarce resources and

reproductive preferences of individuals within a population. In economic systems, on the other hand, customer choices in a competitive market can determine which companies survive and which do not.

In both cases, the evolutionary context establishes conditions which *select for* some units and *select against* others. The context sets criteria which drive the selection processes. The importance of context, then, is paramount. For example, a species that does well in hot climates may fare poorly in colder regions if it has no fur. A car that sells well in North America may not sell at all in Japan because it is left-hand drive. As units evolve in a given context through the processes of innovation and selection, they will embody an evolutionary strategy that meets the criteria of that context. When contexts change, as they often do, so will the criteria. Sometimes a unit must adapt to the changes, other times it need not. An evolutionary unit will prosper in a given context as long as the selection mechanisms driven by that context continue to select for, and not against, that unit.

These four concepts -- unit, context, innovation and selection -- are common to all evolutionary systems. Building on these general concepts to develop evolutionary theories specific to certain fields is appropriate; blindly applying biological processes such as mutation to non-biological systems is not. This thesis will build on these four basic ideas as it looks at the hierarchical evolution of human socio-technical systems.

2.2 Why Study Evolution?

Evolutionary theory emerged to explain the biological change suggested by the fossil record and the diversity of life observed in nature. The theory has since been used to explain social and technological change as well. Evolution, though, is not just a theory of the past, but also of the present and the future. Evolutionary theory helps one to understand not only how we got here, but where we may be heading. It helps one understand the ongoing processes of change in the world.

Understanding change is more important today than ever before. As the human population continues to grow, it pushes the limits of biological and social sustainability. Like many nations, Canada and its provinces are in a debt crisis. The entire world, at least from humanity's perspective, is in an environmental crisis. The rallying cry of governments and activists alike is "we must change." We must change our social systems if Canada is to solve its debt problems. We must change our consumption and production practices if we are to solve the environmental crisis. We must change our very way of life if humanity is to survive.

Others may argue that people are overreacting, that things will work themselves out in the long run. Only time will tell for sure. The choice is between having things work themselves out and working things out for ourselves, between taking a passive or an active role in determining the future of human society. Given that choice, many would elect to take an active role. This thesis is directed towards those individuals. It uses evolutionary theory to help understand change in humans social and technological systems and, by doing so, hopefully provides a tool for those people who elect to take an active role in shaping the future of human society.

2.3 Evolution and Systems

Before looking at the evolution of human socio-technical systems in particular, one should consider at least briefly the evolution of systems in general. What is an evolving system? A system is a whole made up of interdependent parts and existing within a larger context. It was noted above that the definition of an evolutionary unit by itself satisfies two parts of the system definition. It is a whole and it exists within a larger evolutionary context. It was also noted, therefore, that if an evolutionary unit consists of interacting parts then one can call it a system. Evolutionary units are subject to the processes of innovation and selection. An evolving system, therefore, is a whole made up of interdependent parts, existing within a larger evolutionary context and subject to the processes of innovation and selection.

The distinction between systems in general and evolving systems in particular is another restriction this thesis places on its examination of systems. This thesis is interested only in those systems that evolve: those systems existing within a larger evolutionary context, consisting of interdependent components, and subject to the processes of innovation and selection. In Klir's terminology, this is primarily a restriction on things, *T*. The previous chapter already restricted things to humans and the tools they use. This latest restriction refines the earlier one, focusing the study further to investigate humans and tools which evolve.

2.4 Evolution and human socio-technical systems

In human systems, humans use tools and each other to satisfy their needs. Those needs range from simple survival to maintaining a high quality of life, and depend on the context in which the humans find themselves. The tools people use and the socio-technical systems in which they use them have changed significantly over the centuries. Evolutionary theory helps one understand these changes.

Applying evolutionary theory to explain change in human society is not a new idea. Darwin published his On the Origins of Species by Means of Natural Selection in 1859, though in truth evolutionary concepts were not entirely new even then.[4] As early as the 1870s, anthropologists such as Edward Burnett Tylor and Lewis Henry Morgan proposed that modern civilization arose through evolutionary processes.[5] Later, the ideas of Social Darwinism were used to justify a laissez-faire attitude towards areas such as safety regulation and social support policies.

While tools certainly do *not* possess DNA, their development over time does seem to mimic the gradual and sudden changes in biological species observed in the fossil record. One is easily tempted to apply the term evolution to these patterns of development. Is this appropriate? It is if one can show that the four general aspects of evolution outlined above apply to the development of human tools and the socio-technical systems in which humans use them. This section examines each of the four aspects and shows that an evolutionary model is indeed appropriate for modelling change in humans socio-technical systems.

a) What is the unit of evolution in human socio-technical systems?

There is no one answer to this question. Since evolutionary units are systems, defining an evolutionary unit is in the eye of the observer. Different people will identify different units depending on who they are, on why they are defining the unit, and on the context in which they do so. The previous chapter identified the components of human socio-technical systems as humans and their tools. Since one can model both humans and tools as systems, they are both candidates for units of evolution. Humans, of course, are biological species and therefore subject to the process of biological evolution. However, this thesis is interested primarily in *socio-technical* evolution, not in biological evolution. This section argues that humans and their tools together make up the predominant unit of evolution in socio-technical systems.

This is not a new idea. In 1867, one of Darwin's contemporaries, Karl Marx, published Capital in which he suggested Darwin's evolutionary ideas could apply to human technology as well. According to historian George Basalla,

> Marx' evolutionary analogy includes two stages. In the first stage technology engages humanity in a direct, active relationship with nature. Men and women use their labor to shape physical reality, thus creating the artifactual realm. Once the natural world is transformed by work, **nature becomes a virtual extension of the human body** [boldface added]. Thus, men and women working with natural objects and forces bring nature within the sphere of human life.
>
> Having minimized the differences between the made and the living worlds, Marx moves on to the second stage of his argument and suggests that the Darwinian approach to the "history of Nature's Technology" be transferred to the "history of the productive organs of man." He argues that evolutionary explanations should be applied to the organs that plants and animals rely upon for survival *and* to the technological means that humans use to sustain life. Given his assumption that important features of the human body can be explained in evolutionary terms, then so too can technology, **the body's extension into nature**

> [boldface added].[6]

Other thinkers have echoed this idea of technology being an evolutionary extension of man into nature. For example, William Kuhns expresses a similar idea in The Post-Industrial Prophets when discussing Buckminster Fuller's concept of human evolution.

> Man has reached a threshold of evolution where organic changes become transferred to the environment, so that rather than altering the shape of the hand or the accuracy of the eye organically, man hastens his evolution by extending his hand or his eye through tools and instruments.[7]

Both Marx and Fuller considered technology an extension of the human body which is shaped by evolutionary processes much in the same way evolution shaped human hands and eyes. Tools enhance their users' ability to interact with the environment. For example, carpentry tools help people manipulate wood; satellites help people forecast the weather. Humans discovered long ago that tools help them survive and that it was much quicker to adapt tools to changing environmental conditions than evolve themselves. Humans have in effect *externalized* a great deal of their evolution to their tools, whether or not this was a conscious act.

But the tools were not the only thing to adapt to changing environmental conditions. New tools require new knowledge and new skills. Humans, then, have to adapt to the new tools in learning how to use them. Those humans that adopt, and adapt to, the right tools are able to satisfy many of their needs. For example, the careers section of any newspaper lists dozens of career opportunities. Yet if a person does not possess the right tool set, these opportunities are closed to that individual. Those that have the right tool set can find employment and are often better able to satisfy their needs. It is this interaction, this interdependence between humans and

their tools, which suggests using humans and the tools they use, together, as evolutionary units.

Since humans and their tools interact together to form a whole within a larger context, one can model them as systems. As such, they are certainly good candidates for evolutionary units. However, while evolutionary units are systems, not all systems are evolutionary units. To be an evolutionary unit, the system must exist within an evolutionary context and be subject to the processes of innovation and selection.

b) What is the evolutionary context in human socio-technical systems?

What is a good or poor choice of tool set is entirely dependent on the needs of the individual and the context in which that individual seeks to satisfy those needs. Everything exists within a larger context, but not everything exists within an *evolutionary* context. To be an evolutionary context, the context must determine selection criteria for its systems. In socio-technical systems, one can identify many different contexts. The question, then, is do these contexts set selection criteria for the systems they contain? The answer is yes.

One could try to list the many types of socio-technical contexts such as companies, cities and countries, and then try to show that each one sets its own selection criteria. However, since that list is likely endless, it would be a futile attempt. The simple fact of the matter is that humans, as individuals, groups and a species, must work at maintaining both the quantity and quality of their life. How successful they are, relative to their own evaluation and that of others, depends a great deal on the how well the methods they choose to use work in any given

circumstance. It is the context that has the last word on how well something will work, on whether or not someone will satisfy their needs given their particular tool set. The context sets the conditions, the criteria, by which all approaches are judged.

c) Are there Innovation Processes in Human Socio-Technical Systems?

This author certainly assumes they exist. Processes such as engineering design and entrepreneurial inspiration are everyday occurrences in the modern world. If the reader believes otherwise, it is unlikely that any discussion here will alter that opinion. Which processes operate on specific evolutionary units depends on the context in which the unit exists, and on the unit itself.

Assuming such processes exist, it is worth discussing whether they are in response to or in preparation for changes in selection criteria. In the history of *biological* evolution, the latter was first proposed but it is the former than is now generally accepted. In the early nineteenth century, Jean-Baptiste de Lamarck proposed a theory to explain biological change. In it, he supported the even earlier idea of the inheritance of acquired characteristics. This hypothesis suggests that environmental changes *cause* adaptive changes in organisms which they then pass on to their offspring.[8] Innovation, according to Lamarck, is an adaptive response to environmental change.

Darwin, and his contemporary Alfred Russel Wallace, proposed an alternate explanation: natural selection. In natural selection, natural processes select species that are well adapted to

their current environmental context for survival and select others for extinction. The major difference between the two hypotheses is that Lamarck's suggests that existing species change in response to their environment while Darwin's suggests that change *must have already happened* for it to be selected for and not against. These changes are generated internally by what was later discovered to be genetic processes.[9] Natural selection works on an already existing diversity of life while Lamarck's hypothesis lets species diversify after the fact. In both cases, existing species are adapted to their environment. With Lamarck, the environment imposes this adaption externally in *response* to change while, with Darwin and Wallace, the adaption occurs internally in *preparation* for change.

Which is right? Do innovation processes work in response to change or are they ongoing, generating new variations constantly and being selected by as contexts change. In human socio-technical systems, at least, it is a mixture of both. Humans have an advantage over many species in that they can react to changing contexts by adapting their tool sets to compensate. On the other hand, people are also constantly developing new tool sets in the hope that they will work better in existing contexts. Humans can sometimes also anticipate change and adapt *before* change happens. Humans and their tools are perhaps the most flexible evolutionary units in existence because of their ability to innovate in so many different ways.

d) Are there selection processes at work in Human Socio-Technical Systems?

It seems reasonable that humans and their tools are good candidates for evolutionary

units, that they exist within evolutionary contexts, and that they are subject to processes of innovation. What remains is to show that they are also subject to processes of selection.

Selection only works if the process selects for some evolutionary units and against others. In modern socio-technical systems, this is largely a result of the interaction between economics and human needs. Humans have finite resources with which to satisfy their needs. Perhaps the most finite of these resources is time. Biology places strict time constraints on things like acquiring food and shelter. Even if these basic needs are easily met, humans have only a limited life span in which to achieve higher level needs such as self-esteem.

In modern society, satisfying all these needs requires tools. Some tools will work better than others depending on the goals of the individual and the context in which that individual exists. People will select some tools and discard others as they seek to satisfy their needs. As certain tools are selected, others will arise from innovation processes to challenge them, and the cycle continues. People, then, act as selection processes in determining which tools continue to evolve and which become extinct. But how does this translate to the evolutionary unit of humans *and* their tool set? The answer is unfortunately a bit circular.

The previous chapter argues that one can consider humans as tools themselves. The result is the global network of trade that has emerged over the centuries, in which people use each others as tools to satisfy needs they could not satisfy otherwise. People, then, again act as selection processes by choosing to use some people as tools and not others. More accurately, they choose to use some people *and* those people's tool sets. The select for and against the

evolutionary unit of a person and their tools.

Why does it matter if somebody decides to use another person as a tool or not? It matters because, in exchange for being used as a tool to satisfy someone else's needs, that person usually receives goods, services or money which he or she can in turn use to satisfy his or her own needs. If no one uses a person as a tool, then that person will not receive these resources and must find other means, such as unemployment insurance or welfare, to satisfy needs.

This is not to say that other selection processes, such as natural selection, are not present as well. After all, humans are still part of a biologically evolving system. It is only to say that personal choice is the primary selection process in the evolution of socio-technical systems.

Given the above discussion, it seems reasonable to use evolution as a model to explain change in human socio-technical systems. It remains to be seen if it is useful to do so.

2.5 Summary

This chapter shows evolution to be a general theory applicable to many fields by identifying evolutionary units, evolutionary context, innovation and selection as the four fundamental aspects of evolution. One can apply these four evolutionary concepts to explain change in socio-technical systems. This thesis recommends we consider humans and their tool sets together as evolutionary units.

The previous chapter introduced the idea of systems as the first conceptual building block

of this thesis. This chapter discussed evolutionary theory as the second building block. The next chapter examines the concept of hierarchy which is the last major building block of this thesis.

1.. Kenneth L. Feder and Michael Alan Park, Human Antiquity: An Introduction to Physical Anthropology and Archeology (Toronto: Mayfield Publishing Company, 1993), 22.
2.. Kevin Kelly, Out of Control: The Rise of Neo-Biological Civilization (Don Mills, Ontario: Addison-Wesley Publishing Company, 1994), 370.
3.. Kevin Kelly, 373.
4.. Feder and Park, 22-25.
5.. Feder and Park, 30-31.
6.. George Basalla, The Evolution of Technology (New York: Cambridge University Press, 1988), 207.
7.. William Kuhns, The Post-Industrial Prophets (New York: Weybright and Talley, 1971), 236-237.
8.. Feder and Park, 22-23.
9.. Feder and Park, 23-25.

Chapter Three

Concept 3: Hierarchy

3.0 Introduction

This chapter examines the concept of a hierarchy, which is the third and final building block of the thesis. Hierarchies are useful tools for modelling the structure of evolving systems. Chapter One argues that there are at least three levels of description necessary when analyzing systems:

a) the system as a whole;

b) the system as an aggregation of interacting parts; and,

c) the system as a part of a larger context.

Analyzing the system as a whole provides insight into *what* a system does. Analyzing it as an aggregation of interacting provides insight into *how* the system does what it does. Finally, analyzing the system as a part of a larger context provides insight into *why* it does what it does. The result is a three level *hierarchical* description of the system (Figure 7).

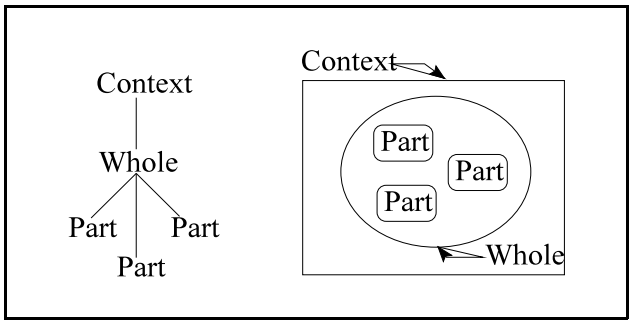

Figure 7: System as Three Level Hierarchy, Non-Nested and Nested Representations

The concept of hierarchy, like system and evolution, needs clarification. That is the purpose of this chapter. It begins by looking at what a hierarchy is and then discusses why one should use hierarchies. Following this, the chapter examines how the concept of hierarchy relates to the concepts of system and evolution which were introduced in earlier chapters. It is intended to suggest that systems have evolved hierarchically.

3.1 What is a Hierarchy?

Hierarchy generally describes a layered representation of a system with a top, a bottom and many levels in between. There are two different ways to classify hierarchies, as nested and non-nested. While both types can be used to describe systems, this section will show that the nested hierarchy best represents the concept of a system.

The traditional conception of a hierarchy is best exemplified by the management tree of a company (Figure 8). The company president resides at the apex of the hierarchy, vice-presidents are placed below the president, then various levels of management follow until the bottom level of employees is reached. This type of hierarchy often implies a flow of control, with the higher level elements controlling the lower level elements, and is called a *non-nested* hierarchy.[1] While a non-nested hierarchy shows some connections between components in a system, there is no sense of the system as a whole made up of interacting parts.

In a non-nested hierarchy, the control structure often implies a sense of higher vs. lower levels. In contrast, nested hierarchies convey a sense of larger vs. smaller levels. In nested hierarchies, the larger levels actually contain the smaller; the smaller levels are *nested* within the larger. In a nested hierarchy, the larger levels cannot, in fact, exist without the smaller

levels. A human body is an excellent example (Figure 9). Very roughly speaking, the human body contains organs which contain cells which contain molecules which contain atoms and so on. It is a nested hierarchy. While the larger levels do *contain* the smaller levels, this does not necessarily imply that the larger levels actually *control* the smaller levels. In fact, a human being may have little control over what his or her heart does while the heart may have a great deal of control over what a person can and cannot do.

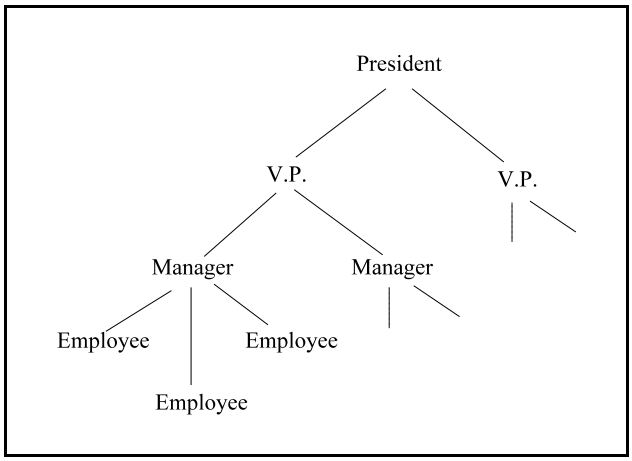

Figure 8: Non-Nested Hierarchy

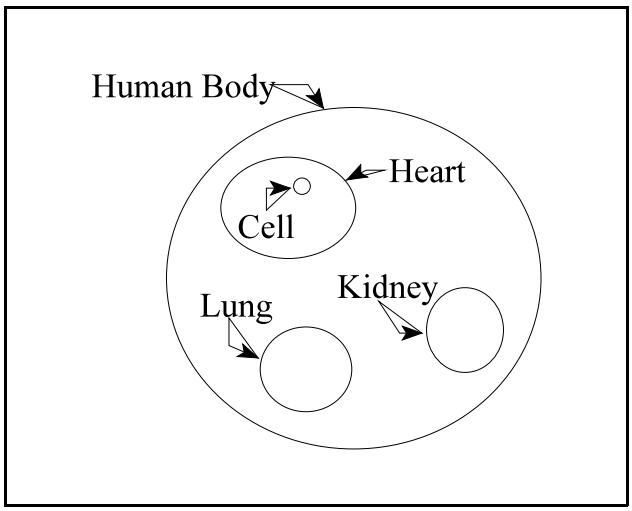

Figure 9: Nested Hierarchy

Most systems are amenable to description by both nested and non-nested hierarchies. However, there are some important distinctions between the two representations. Consider the following two examples (Figure 10).

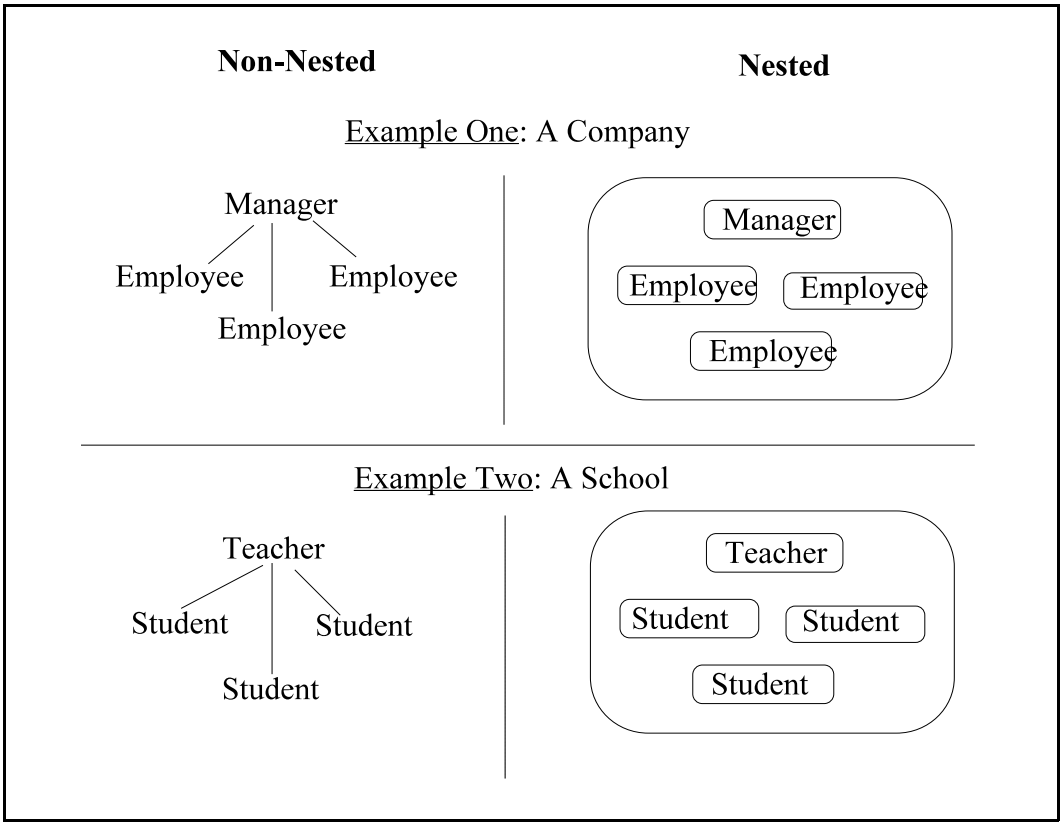

Figure 10: Non-Nested and Nested Representations of the Same Systems

Example One: A company

Using a non-nested hierarchy to describe a company would place a manager one level *higher* than his or her staff. In a nested hierarchical description, the manager and staff are at the same level. The next *larger* level is the project *team*, which includes both the manager and the staff. The non-nested hierarchy may more accurately represents the respective levels of authority in the company, but the nested hierarchy more accurately represents the actual

interactions between staff in the company. The manager may manage the employees, but it is the manager and employees working *together* that creates the greater result of the project team. In other words, the nested hierarchy better represents the concept of a system as a whole made up of interacting parts and operating in a larger context. That larger context is the next larger level in the nested hierarchy. In this case, for example, the team is the context for the manager and employees and the company, in turn, is the context for the team.

Example Two: A school

In a non-nested hierarchical description, a teacher is one level higher than the students; the teacher is in charge of the students. In a nested hierarchical description, the *class* is a whole in which the teacher and the students interact. Again, the nested hierarchy better represents the actual structure of the class as a system made up of interacting parts and operating in a larger context. In this case, that context might be an elementary school.

There is a tendency when representing systems as non-nested, control hierarchies to infer that the higher an element is located in the hierarchy, the more important it is to the operation of that hierarchy. The implication is that because the vice-president has much more authority in the organization than, say, an employee on the assembly line, his or her position is more important than that of the assembly line worker. This is a misconception. Ask any corporate executive who has morale problems on an assembly line. Employees are necessary for the functioning of the company.

Nested hierarchical representations of these same systems clear up this misconception.

Since the larger levels in a nested hierarchy not only contain but actually *consist of* its sub-components, the larger levels cannot, in fact, even exist without their smaller components. The project team cannot exist without its team members just as a human being cannot exist long without his or her heart or lungs.

But while lower level components are *necessary* for the existence of complex systems, they are not in themselves *sufficient*. If 1000 people entered a factory, it is unlikely they would accomplish much unless organized into groups supervised by management. Again, this is shown explicitly in nested hierarchies. Just as the design team would not exist with its employees, nor would it exist without its manager. Once again, it is the manager and employees working *together* that constitute a functioning design team.

The lesson here is that all levels in a hierarchical system are important for its continued operation. The whole may be greater than the sum of its parts, but the whole also needs its parts and the parts need their whole. What is important from a systems theoretical point of view is *how* all the parts interact to form the whole, not just who or what is in charge. The nested hierarchical representation of systems is more amenable to this type of analysis because it represents systems as wholes made up of interacting parts and existing within a larger context. It better reflects the three levels of analysis required to understand a system .

One should not assume, however, that there is only one nested hierarchical description of a given system. Just as the system is defined by the observer, so is the hierarchical description of that system. When creating a nested hierarchical description of a system, one must partition, or separate, the system into hierarchical elements. *How* someone partitions a hierarchy depends a great deal on *who* they are and *why* that system is being analyzed in the

first place. An ecologist looking at a forest might see a complex network of interacting flora and fauna. A developer looking at the same forest might see an opportunity for a great park. Partitioning decisions are largely a matter of personal choice. Two different people may see two different hierarchies when looking at the same system.

Since everyone partitions the world differently, a useful approach is to treat the nested hierarchies of the world as effectively continuous, that is to say they have no fixed boundaries and no distinct levels. People then impose artificial boundaries to partition the *continuous* hierarchy into *discrete* levels to suit their own purposes. However, "discrete levels need to be recognized as convenience, not truth."[2] Where a person places these boundaries of convenience depends on why he or she is describing the hierarchy in the first place. It is important to remember that these boundaries are artificial, placed there for the convenience of the *observer* and not necessarily for the convenience of the *system*.

Consider more closely the earlier example of a company. Where should one really place the boundaries? Surely the four members of the design team are not totally isolated from other employees in the company; surely there are interactions between individuals in different teams. Should one include these other people within the defined boundary? How should one partition the company? There is often no easy answer because, when moving up and down a nested hierarchy, it is generally difficult to define exactly where one level ends and another begins. The hierarchy is effectively continuous. The hierarchy only becomes discontinuous when someone imposes artificial boundaries between *perceived* levels. Looking at nested hierarchies in this way explicitly acknowledges the role of the observer. Each person must make conscious choices about the structure of the hierarchies they observe and describe.

Therefore, each individual is likely to see a different hierarchy.

The relative nature of hierarchical system descriptions can lead to difficulties when communicating these descriptions to others who may perceive different hierarchies. The only way around this difficulty to explicitly state the assumptions used in partitioning the system and to accept the perceived hierarchies of others as valid from *their* perspective as observers. Despite these challenges, there are good reasons for using hierarchical descriptions of systems.

3.2 Why Use Hierarchies?

What is the advantage of using hierarchies to describe systems? Why complicate matters by introducing multi-levelled hierarchies when it might be sufficient to use the three levels of analysis required to define a system, namely as a whole, as an aggregation of interacting parts, and as a part of a larger context ? The answer lies in the nature of the interacting parts and larger context. The whole is not just made up of interacting parts, but of smaller, interacting *sub*systems. These subsystems are in turn made up of even smaller, interacting subsystems. As well, a system is not just part of a larger context, but of a larger *super*system. This supersystem contains many other systems beyond the one of interest, and is itself generally part of an even larger supersystem.

The terms subsystem and supersystem are both relative to the term system. Having identified a system, the larger context is its supersystem and its interacting parts are its subsystems. The relative nature of the terms is important. From the perspective of the subsystem, for example, the system is actually *the* supersystem. And from the perspective of the supersystem, the system is actually one of *its* subsystems. This implies that any system can

actually be a supersystem or a subsystem as well; it just depends on the perspective one uses. Since both supersystems and subsystems are systems in themselves, they share all the properties of a system.

Describing complex systems as hierarchies helps people to understand them by isolating the system into distinct levels. In this way, each level in the hierarchy can be analyzed more or less *independently*. The interactions at higher levels are condensed into the concept of the supersystem, or context. The interactions at the lower levels are aggregated into the subsystems, or parts. This *temporarily* simplifies the whole system. This makes things workable by letting people decide what they need or are able to know. Once a basic understanding of one level is achieved, one can then move up, down or across the system hierarchy to gain a greater understanding of the entire hierarchy.

For example, consider a novice computer user. The novice might begin by learning that the computer system consists of a processor to run software, a keyboard to input data, a mouse to control the user interface, and a monitor to display the results. Once the user is comfortable with the basics, he or she may then choose to delve deeper into some of the subsystems, perhaps learning how to program. Alternatively, the user may decide to investigate some possible supersystems by connecting to the Internet. By initially ignoring the intricacies of each subsystem, the novice can quickly learn how to use a computer to complete basic tasks. As the user becomes more experienced, he or she may learn more about each subsystem and/or more about the supersystems to which it may belong. Once a basic understanding of one level is achieved, the user can move up, down or across the system hierarchy.

Having looked at what hierarchies are and why one should use them in general, the next task is to show how one applies hierarchies to humans socio-technical systems in particular.

3.3 Nested Hierarchies and Human Socio-Technical Systems

The interacting parts of socio-technical systems are humans and their tools. Given the above discussion, then, one should be able to model both humans and tools as systems. One should also be able to model the larger socio-technical contexts as systems as well. Intuitively, this makes sense. Think of any tool, and it is likely to be amenable to description as a nested hierarchy. Electric drills have a casing, a motor, a switch, a plug and a drill bit. Going down the hierarchy, the motor is made of even smaller parts. Going up the hierarchy, one could consider the drill and its operator together as a tool. This larger tool could fit into an even larger system, say a construction firm. A developer could use the firm as a tool to build houses. Moving across the hierarchy at this level, individuals could use the developer as a tool to provide affordable housing. Each level in the nested hierarchy is a system within a larger supersystem and made up of smaller subsystems.

In this example, one begins to see the difficulty of using terms such as larger and smaller to describe relationships between levels in nested hierarchies. If a developer uses a firm as a tool to build houses, this implies they are at the same level in the hierarchy but the firm is certainly larger than a single individual! This thesis introduced the adjectives larger and smaller to differentiate relationships in a nested hierarchy from those in a non-nested hierarchy, where one generally uses adjectives such as higher and lower. Since the terms larger and smaller are *relative* to the particular system in question, be it a person or firm, this thesis will

continue, for lack of suitable alternatives, to use them to describe relationships between levels in nested hierarchies.

Since one can describe socio-technical systems using nested hierarchies, it follows that one should be able to describe *evolving* socio-technical systems using them as well. The next section examines the implications of nested hierarchical structure for evolution, first in general and then for socio-technical systems in particular.

3.4 Nested Hierarchies and Evolving Systems

Earlier sections in this chapter compare nested and non-nested hierarchies. Each has its strengths and weaknesses; each has its appropriate usage. From a system's perspective, the nested hierarchy best represents the part/whole dichotomy of systems. A system is both a part of a supersystem and a whole containing smaller subsystems. A similar dichotomy exists in evolutionary systems where we encounter the unit/context dichotomy.

The pervious chapter argues that if evolutionary units are made up of interacting parts, then they are systems since they are also wholes and exist in larger evolutionary contexts. If an evolutionary unit is a true system, then in fact it is made up of interacting subsystems and exists within a larger supersystem which *acts* as its evolutionary context. If a supersystem can act as evolutionary context, then so can a system if it defines selection criteria for *its* subsystems. Finally, if the evolutionary context is actually a supersystem, then it is also a system and is a candidate for being an evolutionary unit in its *even larger* evolutionary context. For example, the human being as an evolutionary *unit* evolved within the context of the global ecosystem. The human being as an evolutionary *context* sets criteria for its organs. In effect, an evolving

system is a nested hierarchy in which each element, or level, is both an evolutionary unit *and* evolutionary context [Figure 11]. Are there too many ifs? In order for the unit/context dichotomy to hold, each level in the hierarchy must both set selection criteria for its subsystems and be subject to the processes of innovation and selection within its supersystem. However, the evolutionary origins of systems provide some support for the claim that each level in an evolving nested hierarchy is simultaneously both an evolutionary unit and an evolutionary context.

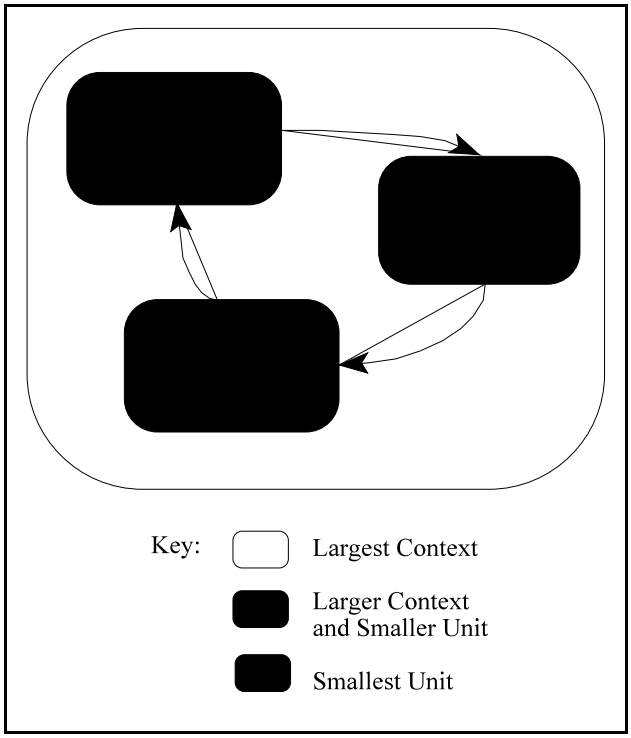

Figure 11: Evolution and Nested Hierarchies

In its definition of a system, this thesis differentiates between the system itself, its context or supersystem, and its components or subsystems. In its definition of evolution, this thesis differentiates between an evolutionary unit and its evolutionary context. Both definitions

imply a hierarchical organization. However, in neither case does the hierarchy require more than three levels. In fact, the evolutionary hierarchy needs only two levels, that of unit and context. Why, then, is it possible to partition most existing evolutionary systems into many more than three levels? Why do most systems seem to exist as systems within systems within systems?

This chapter tries to answer these questions by looking at the origins of systems from an evolutionary perspective. It first presents a model of how complex systems may have emerged over time. It then argues that complex systems which are hierarchically organized would evolve faster than complex systems which are not, and that this advantage explains the predominance of hierarchically organized systems. Finally, it suggests that once a hierarchical system has evolved, the emergence of new intermediate levels *in* that hierarchy is a further advantage, and that this results in the gradual increase of the number of levels in any evolving hierarchy.

3.4.1 The Origins of Systems

Where do systems come from? To answer this question, one must start at the beginning. Systems are made up of interacting components which are able to achieve more by acting interdependently as a whole than they could by acting independently as individuals. The whole is greater than the sum of its parts. However, only in human systems are the parts *aware* of this benefit. In human systems, people can consciously choose to interact. In technological systems, people can consciously design components to interact. However, many systems observed in the world evolved without human intervention. Since they are the precursors of

humans systems, it is important to understand how this happened. Why did the parts of these systems begin interacting if they were not aware of the benefits of doing so?

One can never fully know how any system *actually* evolved if humans were not around to observe it. However, analyzing the properties of systems provides some indications of how they *might have* evolved. These theories can then be applied to existing systems in hopes of better understanding their past and future evolution. One promising proposal is the concept of hypercycles developed by Drs. Eigen and Schuster to explain the chemical and biological origins of the basic building blocks of life on earth.[3] Ervin Laszlo extended this idea to the origins of systems using thermodynamic arguments.[4] This section presents a simplified version of the hypercycle concept.

Most species and systems on the planet transform energy and matter from one form into another. For some, as in the case of plants and animals, this is necessary for survival. For others, as in the case of automotive assembly lines, it is the reason for their existence. This transformation process can be represented as a *flow* through a system of matter and energy. Consider the example shown in Figure 12. To maintain its structure, Component 1 must transform Element A into Element B at a set rate. In order for Component 1 to survive, there must therefore be a readily available supply of Element A. If A becomes scarce, Component 1 will be stressed and will probably perish unless it can find another source of A. One could consider Element B as a waste product in this process.

Now consider another related scenario. Component 1 produces Element B as a waste product. What might happen if Component 1 encountered *another* component which produced Element A as a waste product? Assume that Component 2 transforms Element C into Element

A at the rate required by Component 1. If Component 1 somehow latched onto Component 2, it would have a constant supply of Element A for as long as Component 2 had access to a supply of Element C (Figure 13). In this case, a parasitic relationship is formed between Components 1 and 2 , with Component 1 feeding off of Component 2. This does not necessarily imply that the parasitic relationship was intentional. In the parlance of the previous chapters, Component 1 uses Component 2 as a tool to get Element A.

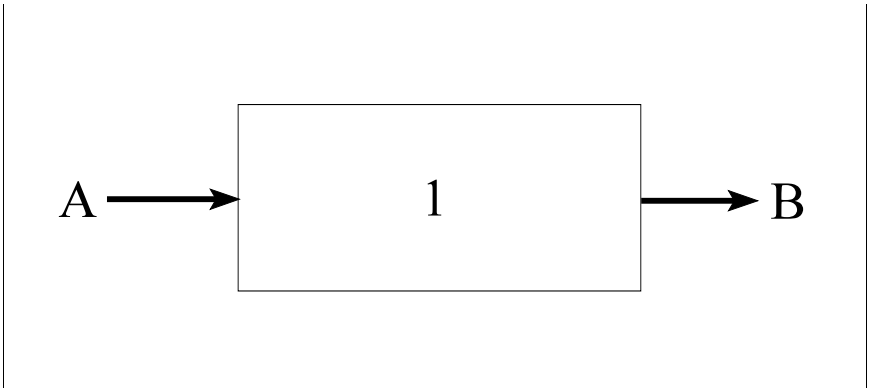

Figure 12: Component One

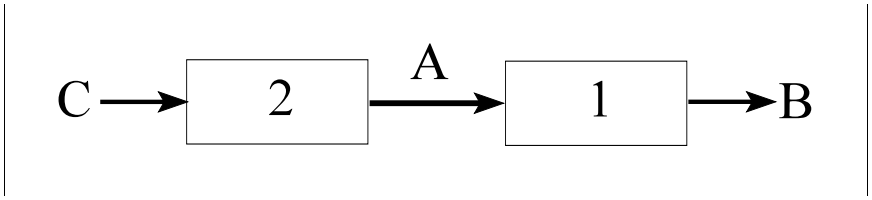

Figure 13: Parasitic Relationship Between Components One and Two

But what if Component 2's input Element was B instead of C. If the transformation rates of the two components were compatible, then the two could enter into a symbiotic relationship instead (Figure 14). In this case, both components use each other as tools to get what they need. Here we have the beginning of a system, in which the parts work together as a

whole.

However, it is unlikely that such a good match would be found. More complicated webs are likely to develop before cycles are formed. The result is the emergence of much more complicated systems with mixtures of parasitic and symbiotic relationships (Figure 15). In other words, networks of trade or exchange. This simple block diagram, for instance, could represent anything from an ecosystem's food web to an automotive assembly line.

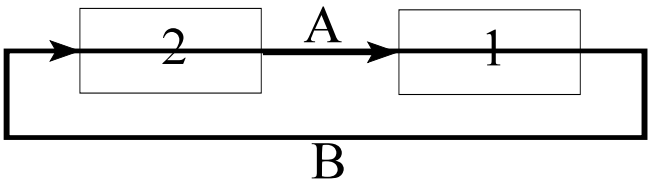

Figure 14: Symbiotic Relationship between Components One and Two

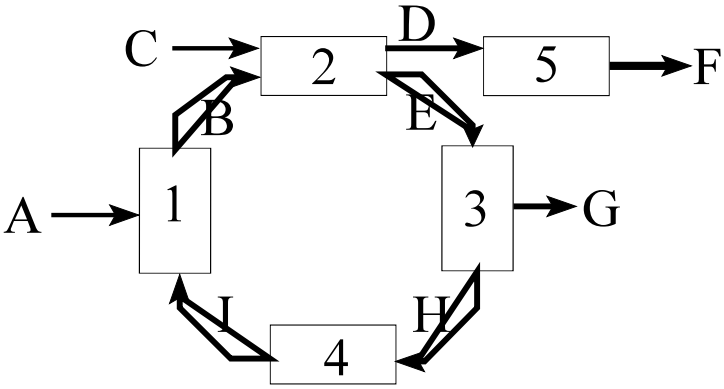

Figure 15: A System of Components

This system transforms Elements A and C into Elements F and G. If this system encountered another component which produced A or C, or needed F or G, then an even larger

system might emerge. As time progressed, it could easily develop into an extremely complicated system as extra components latched on. Consider how this system might look to other components and systems. The system *as a whole* transforms Elements A and C into Elements F and G (Figure 16).

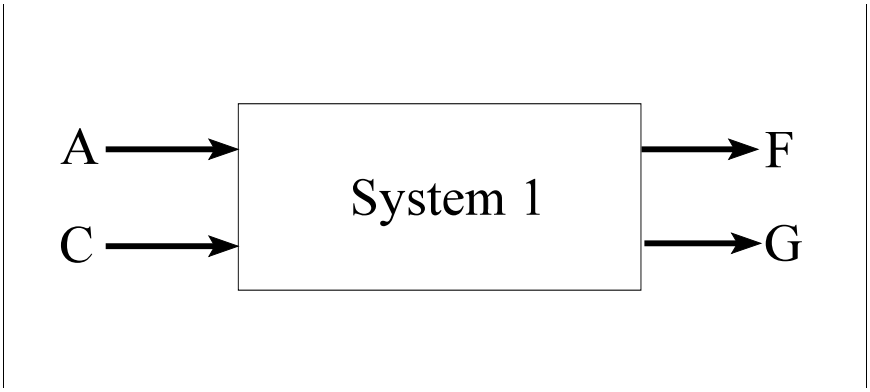

Figure 16: A System as a Component

From the perspective of its environment or context, the internal processes of the system are unimportant. All that matters is the flow *through* the system. In effect, the entire system could be treated as just another *component*, say Component S1. However, it is of little use to do so if new components are added to the existing system one at a time. One might as well keep calling it a system. It is only when two or more previously existing *systems* latch on to one other that it makes sense to refer to some systems as components. When entire systems interact, hierarchies begin to emerge.

However, it is just as easy, and perhaps even easier, for a system to latch onto an additional *component* as it is for it to latch onto an additional *system*. Systems could simply build up one component at a time into hugely complex, but still single level, systems. Hierarchies need not emerge during the process at all. It would seem, however, that they have.

Why? The next section offers one possible explanation.

3.4.2 Building System Hierarchies from the Bottom Up

Herbert Simon, a professor of administration at the Carnegie Institute of Technology, wrote in 1962 that "among possible complex forms, hierarchies are the ones that have time to evolve."[5] He contends that the time required for a system of any sufficiently large number of components to evolve will be much less for a system that evolves hierarchically than for a system that does not. Simon illustrates the logic of this claim with a parable about two watchmakers, Hora and Tempus.

Hora and Tempus both manufacture watches each of which consists of 1000 base components (Figure 17). Hora builds her watch as a nested hierarchy. She first connects the 1000 base components into 100 stable subassemblies of 10 components each. She then uses these 100 subassemblies to build 10 even larger subsystems, each of which now consists of a total of 100 base components. Finally, Hora joins these 10 subsystems together to complete her watch. In total, Hora must build 111 assemblies per watch. Tempus, on the other hand, assembles all 1000 components together at once. He makes only one assembly per watch.

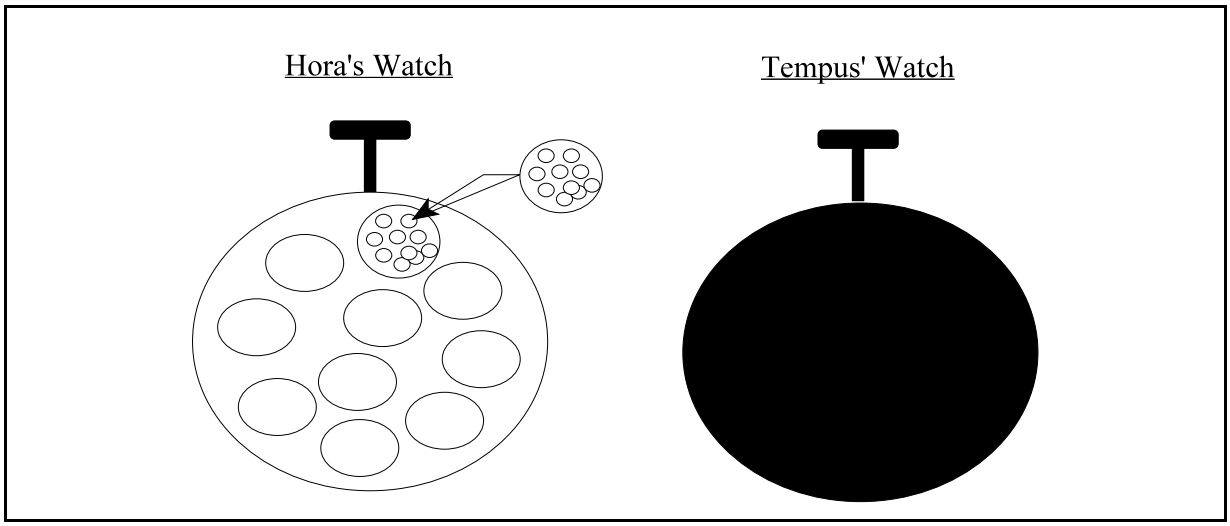

Figure 17: The Watches

Which method will produce more watches? It depends on the context. Assume that it takes the same amount of time, t, to connect a new component to an existing set of connected components, regardless of their size. Under ideal working conditions, it would take Tempus 999t to assemble a single watch as he needs to make 999 connections. In Hora's case, she must make only nine connections per assembly but must build 111 assemblies in total. Therefore, Hora also needs 999t to complete a single watch under ideal working conditions. Tempus and Hora would manufacture the same number of watches in any given period of time if working under ideal conditions.

But what if conditions are not ideal? What if the watchmakers are interrupted at random intervals? Simon assumes that a collection of connected components is only stable when it forms a complete assembly. If either watchmaker is interrupted before completing an assembly, the work put into that assembly is lost. Tempus' watch consists of only one stable assembly of 1000 components. Hora's watch, on the other hand, consists of 111 stable assemblies of 10 components each. When Tempus is interrupted, he could lose a partial

assembly of up to 999 parts. When Hora is interrupted, at most she will lose the work of connecting nine parts. Using stable subassemblies gives Hora a distinct advantage over Tempus under less than ideal conditions.

If the chance either watchmaker is interrupted while connecting a new component to partial assembly is p, with $0 \leq p \leq 1$, then the probability that the component is successfully connected is one minus the chance of failure, or 1-p. For example, if the chance of interruption is 0.25, then the chance of success is 0.75. The probability of making two uninterrupted connections in a row is $(1-p)^2$. The chance of successfully completing an assembly requiring n connections is $(1-p)^n$. The total chance of failing to complete an assembly is one minus the chance of success, or $1- (1-p)^n$. For Hora, n = 9, and for Tempus, n = 999. Nine connections are necessary to assemble 10 components, and 999 connections are necessary to assemble 1000 components. Therefore, the chance of successful connecting 10 pieces together is $(1-p)^9$, while the chance of successfully connecting 1000 components together is $(1-p)^{999}$. The chance of failure is $1-(1-p)^9$ for 10 components and $1-(1-p)^{999}$ for 1000 components. Dividing the chance of failure by the chance of success determines the average number of failed attempts each watchmaker will make before successfully completing an assembly . As shown in Table 1, this ratio varies depending on the value of p. The results in Table 1 show that Hora makes significantly less failed attempts per successfully completed assembly than does Tempus for any value of p greater than zero and less than one.

While Table 1 is useful, it does not indicate how much time, on average, it takes each watchmaker to complete a watch. To begin with, Hora must construct 111 assemblies for every one of Tempus'. Also, the time lost on any given failed attempt to complete a watch is

different for each watchmaker. Only by combining this additional information with the average number of failed attempts per successfully completed assembly can a true comparison between the strategies of each watchmaker be made. Calculating the average time lost per failed attempt is fairly straightforward.

For simplicity, assume that the time, t, required to connect a new component is one second, here represented as s. Uninterrupted, it would take Hora 9s and Tempus 999s to complete an assembly. The average time lost per failure can be computed as follows. Assume that p, the chance of being interrupted while connecting any two components, is 0.25. This means either watchmaker would be interrupted, on average, making one out of every four connections. If interrupted, both watchmakers would on average lose the time to make four connections, which is four seconds. This is the inverse of the chance of interruption. That is, 1/0.25 = 4. In general, the average time lost per interruption is therefore 1/p. But what happens when 1/p is greater than the number of connections necessary to complete an assembly?

Table 1

Number of Failed Attempts per Successful Attempt for Different p

	Hora			Tempus		
p	Chance of Success	Chance of Failure	Number of Failures per Success	Chance of Success	Chance of Failure	Number of Failures per Success
1	0	1	undefined	0	1	undefined
0.99	1.0×10^{-18}	~1	1.0×10^{18}	~0	~1	~undefined
0.75	3.82×10^{-6}	~1	242 000	~0	~1	~undefined
0.5	0.00195	0.998	511	~0	~1	~undefined

0.25	0.0751	0.925	12.3	~0	~1	~undefined
0.12	0.316	0.684	2.16	3.45×10^{-56}	~1	2.90×10^{56}
0.1	0.387	0.613	1.58	1.94×10^{-46}	~1	5.15×10^{45}
0.01	0.914	0.0865	0.0947	4.36×10^{-5}	~1	22 900
0.005	0.956	0.0441	0.0461	0.00669	0.993	149
0.0012	0.989	0.0107	0.0109	0.301	0.669	2.32
0.001	0.991	0.00896	0.00905	0.368	0.632	1.72
0.0001	0.999	9.00×10^{-4}	9.00×10^{-4}	0.905	0.0951	0.105
0	1	0	0	1	0	0

If p is 0.01, then 1/p equals 100. So, on average, each watchmaker will be interrupted making the 100th connection. Tempus must successfully complete 999 connections, so he will lose 100s per failed attempt, on average. However, Hora must successfully complete only 9 connections. If she is only interrupted making the 100th connection, she will successfully construct 11 assemblies, each one requiring nine connections, and then be interrupted attempting the next connection. If this were true every time, then she would lose only 1s per interruption. This is unlikely. A better interpretation is that when Hora is finally interrupted, it will effectively be at a random point somewhere between making the first and ninth connection. She will lose on average (1+9)/2 , or five, seconds per failed attempt. Similarly, when 1/p is greater than 999, Tempus will lose on average (1+999)/2, or 500, seconds per failed attempt.

The average time lost by each watchmaker during a failed attempt at completing an assembly depends on the value of p, but has a fixed value beyond a given threshold (Table 2). The average time lost will increase as p decreases until a threshold is passed for each

watchmaker, at which point the average time lost has a fixed value. However, while the average time lost may *increase* as p decreases, the average number of failed attempts per successfully completed assembly *decreases* as p decreases. If there is a zero chance of interruption, the average time lost is meaningless because the watchmakers are never interrupted.

Table 2 shows that the average time lost during a failed attempt at completing an assembly is generally much smaller for Hora than for Tempus. Table 1 shows that the number of failed attempts per successful attempt is also generally less for Hora than for Tempus. To make a clearer comparison between the two methods, one must combine the results in these two tables to show the average time needed for each watchmaker to complete a single watch. When p is greater than or equal to 0.5, Tempus takes a computationally indefinite number of failed attempts before successfully completing a watch. Hora, at least, completes a watch in a definite period of time. Numeric comparison is therefore only meaningful for values of p less than 0.25.

Table 2

Average Time Lost per Failed Attempt for Different Frequencies of Interruption

	Average Time Lost, in s	
Chance of Interruption, p	Hora	Tempus
1	1	1
0.99	1.01	1.01
0.75	1.33	1.33
0.5	2	2
0.25	4	4

0.12	8.33	8.33
0.1	5	10
0.01	5	100
0.005	5	200
0.0012	5	833
0.001	5	500
0.0001	5	500
0	undefined	undefined

To compute the average time required to complete a single assembly, the average *number* of failed attempts is multiplied by the average time lost *per* failed attempt. This is then added to the time required to successfully complete an assembly, which is 9s for Hora and 999s for Tempus. For Tempus, the result is the average time to complete an entire watch, since he builds only one assembly per watch. For Hora, the result must be multiplied by 111 since she must complete 111 assemblies per single watch [Table 3].

Table 3

Average Time Required to Complete a Watch for Different Frequencies of Interruption

	Hora		Tempus	Ratio of
p	T for 1 assembly	T for 1 watch	T for 1 watch	Tempus over Hora
0.12	27.0	2997	2.88×10^{57}	9.59×10^{53}
0.1	16.9	1880	5.15×10^{48}	2.74×10^{45}
0.01	9.47	1050	2 290 000	2180
0.005	9.23	1020	30 700	30.0

0.0012	9.05	1005	2930	2.92
0.001	9.05	1004	1860	1.85
0.0001	9.00	1000	1050	1.05
0	9s	999s	999s	1

These results show that using smaller, stable subassemblies gives Hora an immense advantage until the frequency of interruptions becomes quite small, relative to the required number of successful connections. In this analogy, the size ratio between Tempus' and Hora's assemblies is a paltry 100:1. As such, the advantage of the smaller subassemblies quickly disappears as the frequency of interruption drops. However, there is still a significant advantage with subassemblies as the frequency of interruptions increases, even for this small a size ratio. The implication for systems containing billions of components, such as living organisms, is staggering.

Extending this simple watchmaker analogy to the evolution of much more complex systems, those systems that are structured as nested hierarchies with stable subassemblies are more likely to evolve than systems that are not hierarchically structured, since the stable subassemblies provide some protection against evolutionary interruptions. The evolution of single-celled organisms greatly improves the chance of multi-celled organisms to evolve if they use the already existing single-celled organism as subassemblies. These subassemblies are the stable intermediary forms which act as *building blocks*. The existence of stable building blocks gives hierarchies time to evolve. According to Simon, this results in the evolutionary predominance of hierarchic systems.

> In such a hierarchy, the time required for systems containing, say, 10^{25} atoms to evolve from systems containing 10^{23} atoms would be the same as the time required for systems containing 10^{3} atoms to evolve from systems containing 10 atoms. The form of the generalization is interesting, in that it describes a relation between two levels of a system that is independent of absolute level.
>
> We conclude that hierarchies will evolve much more rapidly from elementary constituents than will non-hierarchic systems containing the same number of elements. Hence, almost all the very large systems will have hierarchic organization. And this is what we do, in fact, observe in nature.[6]

Simon's argument rests on the fact that it is the number of subassemblies that is important, not the absolute number of atoms. It then follows that complex systems tend to be organized hierarchically because hierarchic systems evolve faster than non-hierarchic systems of the same number of components. There is no way to know this for sure, but the argument has strong logic and an intuitive appeal. It also seems to be supported by many ordinary activities. The time to design and fabricate complex technologies such as airplanes is reduced by treating them as nested hierarchies. These pre-existing subsystems *encapsulate* time, effort and knowledge. They reduce the time, effort and knowledge that must be invested directly as a precondition of success. One need not know exactly how a computer works in order to use a word processor. The user generally lets the computer manufacturer worry about that kind of thing.

However, while Simon's watchmaker analogy is an attractive argument, it is only part of the story. Simon's analogy explains why complex systems may have evolved hierarchically instead of non-hierarchically, but it does not explain why the complex systems formed in the first place. Similarly, the previous section on the origins of systems presented a model of *how* complex systems may have developed, but also did not really explain *why*. The next section of

this chapter attempts to answer why.

3.4.3 Simplifying the System by Complexifying the Hierarchy

Why did complex hierarchical systems evolve. For human systems at least, part of the answer lies in the definition of a system as a whole which is greater than the sum of its parts. In seeking to accomplish greater and greater things, humans have developed complex nested hierarchies which can do more as a whole than any of their parts could do individually. As new systems were developed and adopted, they were used to create even larger systems. This may explain in part why human systems have evolved as nested hierarchies, but it implies someone doing the building. Some systems, most notably the biological, but also the social, seem to emerge for no discernable reason. Why does this happen? Why questions are notoriously difficult to resolve, but a partial answer may be that nested hierarchies simplify the operational environment for all concerned.

Consider a system made up of two components. Assume that there can be no more than one connection between thee two components. That is to say, they can either be connected or not connected. There are therefore only two ways the system could be organized in terms of connections. As the number of components is increased, the number of different ways the components can be organized increases exponentially as the number of components increases arithmetically (Table 4). Consider a social gathering, a party. When the first few guests arrive, the hosts can interact with them easily. The small group can converse together without excluding anyone. But as more an more people arrive, it becomes increasingly difficult to maintain a single group discussion.

Table 4

Number of Connection Possibilities for Systems of Different Numbers of Components

# of Units	# of Possible Connections	# of Connection Possibilities
2	1	$2^1 = 2$
3	3	$2^3 = 8$
4	6	$2^6 = 64$
5	10	$2^{10} = 2048$
n, n>1	$\sum i$, for i=1..(n-1)	$2^{\Sigma i}$

Systems with large numbers of components become quickly complicated if all components are free to interact with each other. A system with only ten components could be arranged in 2^{45}, or $3.5x10^{13}$, different ways. Kelly's massively parallel swarm systems exemplify this situation. To reduce this complexity, there must be restrictions in the number of interactions between components. One way to accomplish this is to build things as clockwork systems, in which the interactions between components are more sequential than parallel.

In the party example, everyone could be arranged in a line. The first person in line could then relate a short anecdote to the second person who in turn could then relate it to the third and so on. In this way, each person is included in the conversation. However, the result in this case would likely be very unsatisfying unless it was part of a party game. A second approach which is more common at parties breaks the party-goers up into smaller groups. Each smaller group maintains its own conversation which can include all the members of the group. In effect, the party is now loosely organized as a nested hierarchy of subsystems.

Consider a system made up of ten elementary components, but structured as a nested

hierarchy with two interacting subsystems of five components each (Figure 18). Assuming that interactions across system boundaries are impossible, or at least rare, the total number of different ways in which this nested hierarchy could be arranged is computed as follows. Each subsystem has five components, and there are 2^{10}, or 2048, different ways in which the components of each subsystem could be connected. Combined, that makes 2048 x 2048 different ways the two subsystems could be organized. The two subsystems themselves can only be connected in one of two ways. Therefore, the total number of possibilities for the system is 2048 x 2048 x 2, or 8.4×10^6.

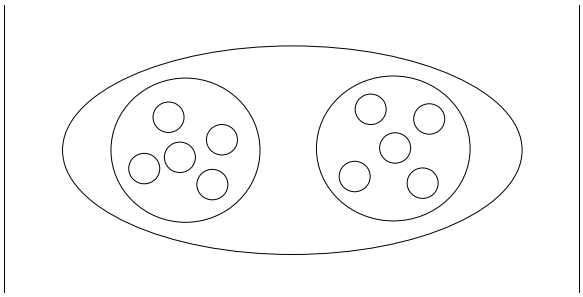

Figure 18: Nested Hierarchy

A *non*-hierarchical system with ten components has 4.2×10^6 times *more* connection possibilities. In terms of how many different ways the components of a system can interact with each other, the nested system is much less complex than the non-hierarchical system. When a party breaks up into smaller groups, each conversation includes less people but can become more focused and coherent.

When systems evolve in the form of nested hierarchies, they simplify the world for their components by acting as a buffer between their components and the system's larger context. As these systems arrange themselves into larger and larger hierarchies, the trend continues. Consider what happens when a new level emerges into an existing nested hierarchy (Figure

19). Emergence here simply means the appearance of something new, in this case a new level.

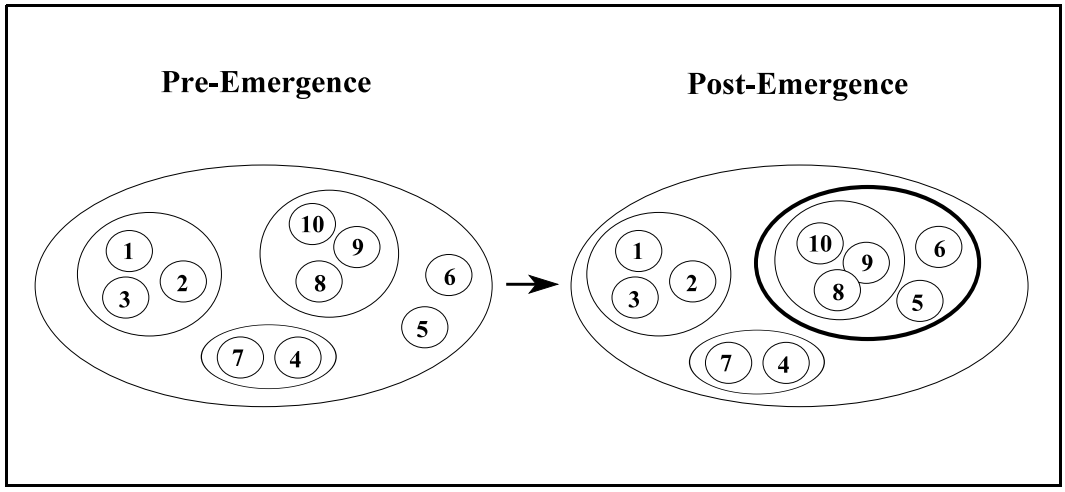

Figure 19: Emergence in Nested Hierarchies

Both systems are nested hierarchies and both have ten elementary components. Before emergence, there were (8 x 8 x 2) x 2048, or roughly 262 000, different ways in which the system could be connected. After emergence, there are [8 x 2 x (8 x 8)] x 8, or roughly 8190, different ways for a total reduction of 32 times. Overall, the entire system has been simplified. The ratio is even greater if only the top level is considered. Before emergence, there were 2048 different ways to arrange the five elements in the upper level. After emergence the number of elements is reduced to three and there are only eight ways to arrange these. This is a reduction of 256 times.

The operational advantage is tremendous. By eliminating unnecessary interactions, components can focus on more important tasks. Consider an employee of a large firm. Under normal conditions, that employee is not concerned with the vagaries of the economy because the company acts as an intermediary. The company pays the employee a regular wage so that the employee can concentrate on doing his or her work. The company president, in turn, *can*

concentrate on the vagaries of the economy because the employees do *their* jobs. Nested hierarchies simplify complex systems for the participants by i*solating* components from one another.

However, this isolation has disadvantages. For example, changes at one level may surprise components several levels removed from the source of change. This can result from, among other things, communication problems across levels in the hierarchy. Employees who let the president worry about the company organization may be caught totally by surprise by a job loss due to corporate re-engineering. On the other hand, the president may not find out about design delays until after committing the company to production deadlines. Chapter Five examines in more detail the implications of isolation. For many systems, though, the benefits must outweigh the costs because emergence happens regularly.

If reducing the number of interactions with the larger environment also reduces the frequency of interruptions that system encounters, this is an added advantage. It was shown earlier that even using stable subassemblies, Hora would still need to make on average 1.0×10^{18} attempts before completing a single assembly with $p = 0.99$. Any system structure that could reduce the frequency of interruption would be a great boon to the evolutionary process. Those systems that exist today may do so in part because the evolutionary units they contain have simplified their evolutionary contexts through nested hierarchies.

To summarize, while describing systems in terms of nested hierarchies may be a human activity, this does not preclude the possibility that the systems of the world are in fact structured hierarchically. Taken together, the last three sections of this chapter suggest that

systems may have evolved hierarchically long before humans were around to observe it. Arthur Koestler, whose works introduced this author to nested hierarchies, goes so far as to suggest that "wherever there is life, it must be hierarchically organised."[7] While one can never prove that hierarchies exist independently of human observation, it is certain that "somewhere between the world behind our observations and human understanding, hierarchies enter into the scheme of things."[8] However, this *does not* imply that the hierarchies people *perceive* exist independently of their observation. Using nested hierarchical *systems* to look at a nested hierarchical *world* does not guarantee that one will see what is actually there. The hierarchies people use to describe the world are mental artifacts. Recall that Chapter One identified mental artifacts as a type of human tool. If used properly, hierarchies are useful tools for understanding the evolution of systems.

If one accepts nested hierarchical evolution as a good model in general, then it makes sense to apply it to human socio-technical systems in particular. In doing so, one begins to observe patterns of development. Before going further, then, one should take a closer look at these patterns. That is left to the next chapter.

3.5 Summary

Hierarchies are useful tools for analyzing the structure of systems. Nested and non-nested are two major types of hierarchies. Of the two, nested hierarchies best represent the concept of a system as an aggregation of interacting parts and as a part of a larger context. In hierarchies, contexts become supersystems and parts become subsystems. Hierarchies help people understand systems by letting them isolate specific levels for analysis.

Artificial boundaries are used to partition effectively continuous hierarchies into discrete levels. How different people partition systems will depend on who they are, why they are analyzing the system, and the context in which they do so. Partitioning assumptions are not always obvious and must be made explicit if they are to be used as tools for communication. Failure to understand the partitioning of a given situation could lead to serious misunderstanding.

Nested hierarchies are also useful for analyzing the structure of evolving systems. In nested hierarchical evolution, each element in the hierarchy is both an evolutionary unit and an evolutionary context. This is the result of the hierarchical nature of systems evolution. Applying hierarchical evolution as a model to describe change in socio-technical systems reveals patterns of development. The next chapter discusses these patterns.

1. T.F.H. Allen and Thomas B. Starr, Hierarchy: Perspectives for Ecological Complexity (Chicago: University of Chicago Press, 1982), 38.

2. Allen and Starr, 6.

3. M. Eigen. and P. Schuster, The Hypercycle: A Principle of Natural Self-Organization (Berlin: Springer-Verlag, 1979), 2-6.

4. Ervin Laszlo, Evolution: The Grand Synthesis (Boston: Shambhala New Science Library, 1987), 28-35.

5. Herbert A. Simon, "The Architecture of Complexity," Proceedings of the American Philosophical Society 106 (December 1962): 473.

6. Herbert A. Simon, "The Organization of Complex Systems," in Hierarchy Theory: The Challenge of Complex Systems, ed. Howard H. Pattee (New York: George Braziller, 1972), 8.

7. Arthur Koestler, The Ghost in the Machine (Hutchinson & Co, 1967; reprint, London: Arkana Books, 1989), 47.

8. Allen and Starr, 6.

Chapter Four

Patterns of Nested Hierarchical Systems Evolution

4.0 Introduction

Testing the usefulness of evolving nested hierarchical systems as a descriptive model of socio-technical change requires the analysis of a real case. In 1973, Pertti J. Pelto published a book on the effect the introduction of the snowmobile had on the Skolt Lapp, reindeer-herding society of northeastern Finland.

> Because the advent of the snowmobile is so new and so stark, we have the opportunity to examine the complex of related technological, economic, and environmental elements that brought about the transition to modern, cash-oriented lifeways. Instead of merely wondering what happened, we are close enough in terms of time to be able to study and to understand how the change actually occurred. Since all this has taken place in the past seven or eight years, what happened is so clearly etched in the increasingly anxious minds of the Skolt Lapps of Sevettijärvi that we can read the printout of the statistical changes *and* recapture some of the play-by-play course of events that has brought these people to the present stage of unremitting sociocultural change.[1]

His analysis makes an excellent case study with which to test the ideas presented in this thesis because he examines the "adaptive responses of different individuals to the new technical and economic situation brought about by the machines."[2]

This chapter translates Pelto's observations into the framework of evolving nested hierarchical systems. It begins by providing background information on the Skolt Lapps and describing their herding practices before the introduction of the snowmobile. It then examines the effect the snowmobile had on this system. The chapter closes by extending the *specific*

changes observed in Skolt society to *general* patterns of structural, evolutionary change in humans socio-technical systems.

4.1 Background Information

This section provides background information on Skolt society by discussing who the Skolt Lapps are, looking at how they lived before the advent of the snowmobile, and commenting on Arctic transportation in general.

Who are the Skolt Lapps?

The Skolt Lapps make an interesting case study for many reasons. When Finland won its independence after World War I, the region in which the Skolts lived switched from Russian to Finnish control and the Skolts became Finnish citizens. In World War II, the Skolts fought with the Finnish army against the Russians. The area in which they lived became a war zone and the Skolts were evacuated in 1940. After the war, the USSR reclaimed the Skolt homeland but the Skolts elected to stay in Finland. The Finnish government eventually resettled the now homeless Skolts in the Arctic Sevettijärvi region of northeastern Finland. Because they lost most of their possessions during the evacuation, the 50 Skolt families started their new lives on roughly equal footing.

How Did They Live before the Snowmobile?

Prior to 1960, the Skolt economy relied primarily on fish and reindeer even though, as Finnish citizens, they receive "free or inexpensive medical services, free schooling, family

allowances, pensions, old age assistance, and other services."[3] Both fish and reindeer were food sources, but the reindeer were much more than that.

> Draught and pack reindeer were a principal means of transportation...; and reindeer hides were important for making shoes, leggings, fur coats, and other articles of clothing. Sinews were used for sewing, and a few antler and bone items were manufactured. Also, reindeer meat, in relatively large quantities, was sold to traders and meat buyers for the cash needed to buy modern necessities, including flour, sugar, tea, coffee, and other staple foods, as well as many nonfood commodities. Thus the reindeer industry was important in a number of different aspects of their mixed subsistence-and-cash economy. As long as a man had reindeer herds he could always sell animals to get needed cash.[4]

In the parlance of this thesis the reindeer were tools, both as living systems and as separate parts.

Since reindeer were so important to Skolt society, most of the Skolt men's winter activities centred on reindeer herding and related activities.[5] Each Skolt family owned at least a small herd of reindeer and maintained herds near there homes throughout the winter and spring. In the summer, however, the reindeer were left to roam free and the men often sought wage labour elsewhere. Ear-markings differentiated one family's reindeer from another, much like the branding of cattle.

A Note on Arctic Travel

Prior to 1960, Arctic inhabitants of Northern Eurasia relied primarily on traditional, non-motorized methods of transportation such as skis and reindeer sleds for winter mobility. These technologies have been around for centuries. Pelto remarks that "archaeological evidence shows that skis were in use in northern Europe at least two to three thousand years

ago."[6] Even though modern, gasoline-powered vehicles were available, the combined factors of high cost, difficult geography and poor infrastructure discouraged their use. Cars, trucks and airplanes require large, up-front investment and are expensive to run and maintain. In addition, rough terrain and the lack of available roads restrict the use of the ground vehicles. In the context of pre-1960s Arctic conditions, the Skolt Lapps largely selected skis and reindeer sleds over motorized transport for winter travel.

This all changed with the invention of the snowmobile. The first snowmobile apparently arrived in Finland in either late 1961 or early 1962, purchased by a school teacher. The school teacher originally bought the snowmobile for recreational purposes, but soon found it practical for hauling wood and supplies as well. News spread and the snowmobile began its infiltration of Finnish and Skolt society.

4.2 Applying the Model

The snowmobile affected many aspects of Skolt society, but this section focuses on their reindeer herding practices. Since 1898, politically created reindeer associations have overseen reindeer herding in Finnish Lapland.

> Every [reindeer] owner in a particular district *must* participate in the association by paying "head" taxes to the association and managing his animals in conformity with the general association (and Finnish) regulations. Since it would be physically impossible for an individual owner to keep his animals totally separated from other animals of the district, the herdsmen's fortunes are inevitably interwoven with the policies and practices of the association.[7]

By law, then, the herdsmen must use the associations as tools to gather their herds. These associations hire herders to round up the reindeer in the fall and winter months. This section

compares herding practices before and after the use of the snowmobiles using modelling framework of nested hierarchical evolving systems.

4.2.1 Herding before the Snowmobile (1955-1960)

Before the snowmobile, the herders all had roughly equivalent status and were paid a daily wage for their work, with a bonus if the herder owned and used a reindeer dog as well. The process was rather lengthy. Herding activities began in the fall and continued throughout the winter and spring. There are seven phases to the process, starting with the capture of the draught animals needed to begin the roundup and ending with the spring calving of new young and the release of the herd until the following fall. Owners let their reindeer roam freely in the summer months.

Phase One: Acquiring Draught Animals

In managing their herds, owners usually castrated some bulls to encourage selective breeding. These geldings were relatively tame and ideal for draught animals. Since reindeer-drawn sleds were an important component of winter transportation, the first stage of the roundup began when herdsmen went out in the early fall to capture some geldings. Herdsmen now had two transportation tools available for winter travel: skis and reindeer-drawn sleds.

Phase Two: Gathering the Herd

As soon as travel conditions allowed in November, groups of three to four men under the informal direction of a foreman would begin gathering the smaller, free-roaming herds into

a single, large herd. This herd was carefully watched to keep it from scattering and to protect it from predators.

Phase Three: Staging the Roundup

Once they collected a large herd of usually several thousand reindeer, local newspapers announced that a roundup would be held at one of the permanent corrals the associations maintained. A herder on skies or riding a sled would lead a bell-reindeer the forty or so kilometres to a staging area. The herd would follow in a long, slow-moving file, stretching out for a kilometre or more behind. Other herders on skis and sleds flanked the file to keep order, sometimes with the help of reindeer dogs.

Phase Four: Corralling the Reindeer

Once at the staging area, they used the same process to lead the reindeer in smaller groups to the corral (Figure 20). A funnel fence guided the reindeer towards this corral. At this stage, caution was required as the animals could easily bolt if startled. The herder with the bell-reindeer quickened his pace and led the reindeer down the funnel towards the corral. Once the last reindeer cleared the mouth of the funnel, "dozens of herdsmen hiding in the snow jumped up and ran into position at the entrance -- closing the funnel with a human fence until the animals were secured in the secondary corral."[8]

Please note that the figure is not to scale. Generally the herders guided several hundred reindeer at a time into the corral.

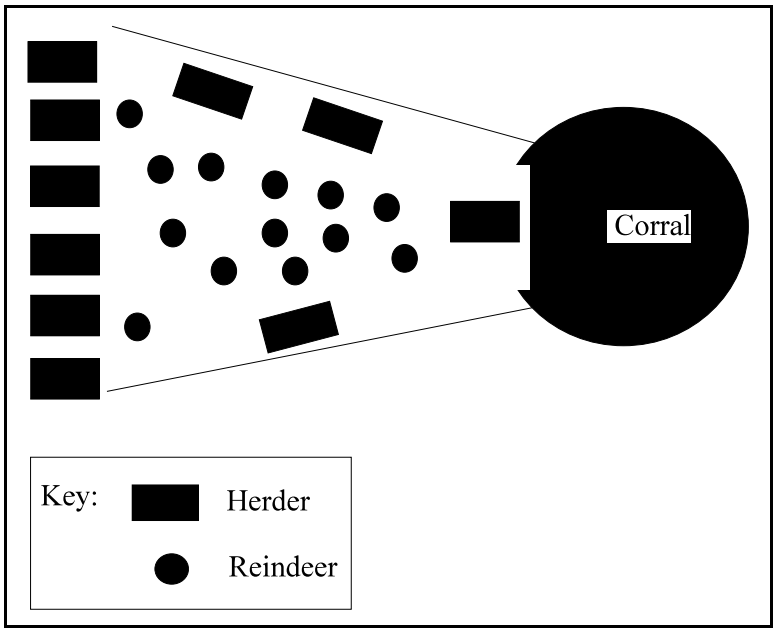

Figure 20: Herding Reindeer into a Corral

Phase Five: The Roundup

Once in the corral, the individual owners could then capture their animals using lassos. They then pushed and dragged their animals to smaller enclosures to separate the herd. At this point, they slaughtered some animals for home use and sale. The young of owned mothers were earmarked as the rest were gathered for tending. Small groups of individual owners would then jointly drive their animals to winter grazing locations near their homes.

The roundup could last anywhere from three days to two weeks. The men would stay in cabins near the corral. Some of these cabins were coffee houses, the proprietors of which would sell hot meals and alcohol.

Phase Six: Winter Herding

During winter herding,

> Owners visited their herds periodically to carry out various tasks such as butchering, castrating bulls, and selecting geldings for training as draught animals. Often many days elapsed before the herd was visited again. These winter herds constituted the food storage and "cash reserve" system of the herders. Whenever supplies ran short, a few animals could be slaughtered for home use and for trading to the stores in Norway.[9]

Phase Seven: Spring Calving and Release

Pregnant cows were grouped together and their young tended to when born. This helped ensure their survival by providing early protection from predators and extra care for sickly calves. Additionally, the owners would earmark all young born to captive cows. Once a few days old and healthy, the cows and calves were released. By early June, the entire herd would be roaming free until the following fall.

4.2.2 The Nested Hierarchical Socio-Technical System

The chapter will now analyze phase four, the corralling of reindeer, as a socio-technical system. There are several possible approaches. One could start by identifying the larger context, and work down: the top-down approach. Alternatively, one could identify all the component parts, and work up: the bottom-up approach. This thesis favours a slightly less rigid approach: the middle-out approach.

Since all systems are in effect nested hierarchies of systems within systems within systems, identifying a fixed top or bottom layer places unnecessary restrictions on the system description. The middle-out approach assumes no definitive top or bottom layer in the hierarchy. Rather, it begins by placing the system of interest in a *middle* layer of the hierarchy

and works its way *out* by moving up, down and across the hierarchy, incorporating additional layers as required. Since socio-technical systems are often effectively unbounded below and above, this method allows the modeller to grow the model out from an anchoring centre point.

Step One: The System of Interest

Figure 20, above, has already identified the system of interest. Viewing it as a nested hierarchy, the system has several components (Figure 21). These are:

1) the reindeer herd;

2) the corral;

3) the human wall;

4) the lead herder;

5) the flanking herders; and,

6) the landscape.

There roles for each of these components in the process are described above.

Identifying these six components implies a system boundary. For example, the rest cabins are not included. Even through the rest cabins are an important part of the whole reindeer herding practice, this section is interested only in those components directly involved in herding the reindeer into the corral. The boundary becomes even more defined when the context is identified in a subsequent section.

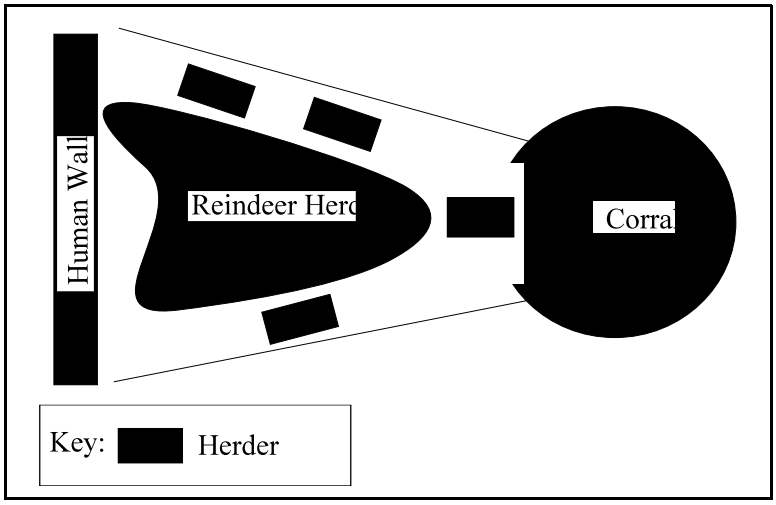

Figure 21: The System of Interest

From here, one could move down the hierarchy, analyzing each of the components as a system, up the hierarchy identifying the nested contexts of which it is a part, or across the hierarchy identifying other systems that share this level. This thesis elects to move down the hierarchy first.

Step 2: Moving Down the Hierarchy

If the above system is truly only one level in a nested hierarchy, then one should be able to analyze each component as a system as well. Consider them in order.

1) The Reindeer Herd

The reindeer herd is a system made up of individual reindeer -- bulls, cows and calves. Moving further down the hierarchy, one could model each individual reindeer as a system made up of cells, and so on. Alternatively, one could try to identify smaller groups of animals within the larger herd.

2) The Corral

As represented in Figure 21, the corral is a fairly simplistic system made up of two components -- the enclosure and the funnel fence. As mentioned earlier, however, there are also many smaller enclosures attached to the main enclosures. Moving down the hierarchy further, the enclosure and fence are both systems of assembled materials such as wood and nails. Both wood and nails are systems are large numbers of cells and molecules, respectively. The process can continue further.

3) The Human Wall

The human wall used to block the entrance of the funnel fence is an interesting system. It consists of dozens of individual herders working together. Each herder is himself a system consisting of a Skolt Lapp and his tools. For instance, as the herd is being led into the funnel, the herder uses snow as a tool to hide from the herd and clothes to keep warm while doing so. All this is possible because they use their physical attributes and mental skills as tools. When making the wall, the herders use each other as tools to construct the wall. One could further analyze the Skolt in terms of his biology and the snow in terms of its composition.

4) The Lead Herder

The group uses the lead herder as a tool to lead the reindeer into the corral area. To accomplish this, the herder uses a variety of tools. Together, the herder and his tools make up a system. In further analyzing this subsystem, one could identify two major components -- the mobile Skolt Lapp and the bell-reindeer he uses to lure the herd. While Pelto never clearly

states so, this thesis will assume for illustrative purposes that the bell-reindeer is a system made up of a reindeer and a some sort of bell. The mobile Skolt Lapp is actually a system consisting of the human being with physical and mental skills and attributes, his clothes and his method of transportation. The method of transportation could be either skis or a reindeer-drawn sled. One could further analyze each of these as systems.

5) The Flanking Herders

The Lapps who flank the herd are also systems consisting of humans and their tools. In this case, the tools are the physical and mental skills and attributes of the Lapp, his clothing, skis or a reindeer-drawn sled, and sometimes a reindeer dog. One can, again, analyze each of these as systems in their own right.

6) The Landscape

The landscape is an important part of the system. One might assume that the natural environment is actually the context and not part of the system. However, this thesis prefers to treat the landscape around the corral as *part* of the system for herding reindeer. The herders depend on the snow both to hide the human wall and to slow the reindeer's movements. But if the landscape is not the context, then what is? The next section addresses that question.

While hardly an exhaustive analysis of the system of interest, the above should demonstrate the nested hierarchical nature of the system as one moves *down* the hierarchy. However, down is only one direction. One could also move up and across the hierarchy.

Pelto's book does not provide enough information to move across the hierarchy in any useful way. Therefore, the next section looks at moving up, at identifying the larger contexts/supersystems in which the system of interest is nested.

Step 3: Identifying Contexts

What is the context in which the herding team operates? The quick answer might be northeastern Finland and all the social and ecological trappings that includes, but this is too broad an answer to be immediately useful. If contexts do indeed determine selection criteria for their evolutionary units, then in attempting to identify that context one could ask what determines the continued existence of the reindeer herding team? Since the herders are in the employ of the local reindeer association, the association does. As long as the association oversees reindeer herding operations, it also oversees the herding team. The reindeer association, then, is the larger context for the team as a system, at least from a socio-technical evolutionary perspective. The association, in turn, exists within the larger context of the northeastern Finland's society and ecology.

It is within the context of the association that changes to herding practices are selected for or against, just as it is within the smaller contexts of herder and his tools, as a system, that the tools are selected. This is evidenced by the changes resulting from the introduction of snowmobiles to reindeer herding in the early 1960s.

4.2.3 The Introduction of Snowmobiles

The snowmobile greatly changed transportation for the Skolts. Trips that once took

days now took a matter of hours. Within a decade, the snowmobile completely replaced skis and the reindeer-drawn sled as the primary means of winter travel. These changes occurred in reindeer herding as well.

> Herders all over Lapland were quick to mechanize their herding operations. Evidently, many herdsmen felt that this modernization would make herding physically easier and economically more advantageous.[10]

The snowmobile was there to stay, but what effect did this have on reindeer herding? One can answer this question by examining the nested hierarchical systems affected.

> Initially, the herders were only interested in a superior method of transportation.
>
> When the herdsmen... first started using snowmobiles, their intention was simply to substitute a much swifter and easier means of transportation for the traditional, tediously slow, reindeer sleds and skis. They had no intention of changing the basic organization of reindeer herding.[11]

The individual herders selected the snowmobile over their sleds and skis using several criteria, the biggest one being speed. Pelto notes several other factors as well. For example, preparing for a trip by sled can take between one and three hours while by snowmobile it only takes five to twenty minutes. In addition, reindeers require rest on long journeys while snowmobiles simply require refuelling. By far, however, it is speed that is the main criteria. A supply trip to Norway that once took 3 days could now be done in 5 hours. Pride and prestige were also factors as people without snowmobiles were forced to hire those that did for transportation.

At this stage, it is at the level of a Skolt herder and his tools that the change occurred. The immediate effect was the displacement of reindeer-drawn sleds and skis as the main method of transportation. In terms of reindeer herding, this eliminated entirely the need for Phase One as the herders no longer needed to capture gelded bulls to pull their sleds. As the

herder and snowmobile integrated together as a system, additional changes resulted to both the herder and the snowmobile.

First of all, the herder's skill set changed as the herder adapted to the new tool. For example, before snowmobiles, herders were skilled at training and driving with draught reindeer, building and maintaining equipment such as harnesses, all-day skiing, and training and using reindeer dogs. Because of the snowmobile, herders became skilled instead in maneuvering snowmobiles, maintaining and doing field repairs on snowmobiles, and coordinating herding operations though visual cues since they could not communicate by voice over the sound of their engines. Some of these changes required expanding the herder's tool set further , most notably in the addition of specialized tools for repairing snowmobiles.

The herder was not the only element of this new system to change, however. The snowmobile also underwent modifications. The first snowmobiles introduced to Finland were designed primarily for recreational use. The Skolts pushed the limits of these early machines. Later machines, incorporating more powerful engines, were preferred by the herders. Changes in design were not the result of herder demands, though. In fact, "snowmobile design appears largely insensitive to the kinds of uses that Lapps find to be most important."[12] The herder population of Finland is, after all, relatively small in a global sense. However, snowmobiles that do incorporate design features preferred by the Lapps tend to sell better in northern Finland. As well, the herders are not above doing their own modifications. One herder actually built his own snowmobile from parts of other snowmobiles and an old car. A major innovation he incorporated was replacing the twin skis then standard on the front end of snowmobiles with a single, wider ski more suitable for driving in his climate.

As the snowmobile and herder co-evolved as a system, changes were also occurring at other levels in the hierarchy. For example, reindeer reacted differently to the noisy and quick moving snowmobiles than they had to the quiet, slow moving ski-men and sleds.

> In pre-snowmobile days gathering the herds was a leisurely process, often lasting several weeks. As larger and larger numbers of animals were accumulated, a social system developed in which the animals became accustomed to one another and to the men and their dogs. Most of the interaction between the men and their animals was relatively peaceful, and the herdsmen even aided the reindeer in the winter food quest by searching out new grazing areas in the deep snow.
>
> This picture of "peaceful coexistence" was abruptly altered when snowmobiles came into extensive use. The tone of herd control changed -- the animals were frequently chased, sometimes for rather long distances, and contacts between men and animals were lessened since the men no longer spent weeks and months in gathering and living with the herds. Partly because of the high costs of motor operation, they sought to sweep the herds together in a few days, transporting them toward roundup sites with a minimum of delay. The quickly gathered herds were no longer in any sense "social systems," but simply bunches of frightened animals forced to run in the same direction.[13]

This change in procedure was selected for by the context. Before the advent of the snowmobile, herders were paid a daily wage by the association with a small bonus if they owned a reindeer dog. Now, the group of herders was paid a flat sum based on how many reindeer they managed to herd into the corrals. This encouraged the herders to both finish the job faster and with less men. Notice here that the association does not itself select for the process, but instead creates a context in which the herders themselves select the process.

There were other changes as well. Since the reindeer were effectively chased to the roundup, the lead herder and his bell-reindeer were no longer necessary, or even useful. Gathering very large herds before driving them to roundup was eventually dropped from the process and temporary burlap corrals were erected to do smaller, more localized roundups.

Instead of traditional funnel fences, lightweight nylon cords with plastic ribbons tied every couple of metres were stretched out from the corral in a V for several kilometres. These acted as psychological fencing for the reindeer. These cords were also added to the existing fencing at the permanent corrals.

Because the animals were so skittish and the new methods gathered few of any individual owners reindeer at any one time, winter herding was eventually abandoned. This eliminated Phases Six and Seven of the herding process. The marking of new born and the selling of the increasing number of unmarked adult animals was carried out during the roundups. These changes had detrimental effects on the social organization of the Skolts. Only a decade after the first introduction of snowmobiles, the total number of reindeer owned by Skolts in the association dropped "from a high of over 2600 in 1960-61 to below 1700 in the spring of 1971."[14] More significant, one family owned roughly a third of these reindeer while most of the other families had suffered serious loses. About two thirds of the Skolt families originally in the reindeer association were effectively eliminated from reindeer keeping. Many of the herdsmen of 1960 had moved on to wage-labour employment or were unemployed.

The Skolts felt they were losing more and more control of their herds and herding practices to other members of their association. In 1961, the Finnish government granted the Skolts an autonomous reindeer district, dividing the original association in two. But while the herders recognized the new political boundary separating associations, the reindeer did not. As a result, at the time of Pelto's book both associations were considering erecting reindeer fences to physically separate the regions. Whether or not these measures helped was not known at the time Pelto wrote his book.

How many of the above changes were the direct result of the snowmobile is debatable. It certainly seems, however, that the snowmobile precipitated a great deal of change when it was introduced to the reindeer-herding population of northeastern Finland in the early 1960s. An exhaustive analysis of the situation was not possible in Pelto's book and is certainly not possible here. The above case does, though, illustrate some of the ideas presented so far in this thesis. The next section of this chapter uses this specific illustrative example of nested hierarchical systems evolution to identify some more general patterns of change.

4.3 General Nested Hierarchical Systems Evolution

This thesis claims that every evolutionary unit exists in a larger evolutionary context, that it is within these contexts that units are evaluated based on selection criteria, and that these selection criteria determine which evolutionary units survive and which do not. It also claims that every system is but one level in a complex nested hierarchy and therefore exists simultaneously as both an evolutionary unit and an evolutionary context. This dual nature of evolving systems has many implications for the evolutionary process since any *single* change to a system will affect *both* its aspects. This section first examines the general phases of the evolutionary process and then discusses their implications.

4.3.1 The Process of Hierarchical Evolution

The process has three major phases. Phase One is the emergence of a new evolutionary unit into an already existing evolutionary context. Phase Two is the refinement of the original unit as it integrates itself into the context. Phase Three is the eventual displacement or

extinction of that unit within that context. The following discussion relates specifically to *human* systems only and not necessarily to systems in general. However, one could extend most of the material to systems in general.

Phase One: Emergence

New systems emerge when their component parts first come together to interact as a coherent whole. A new system can emerge spontaneously, as in the case of the biological origins of human beings. However, most human systems are really tools. As such, if they were not outright designed and constructed, they were at least encouraged to emerge. The usual impetus behind the emergence of a new human system is the ability to do more with the system than could be done before. However, this does *not* imply that the new systems arose out of necessity. George Basalla suggests that the need for many modern inventions, most notably the car and truck, arose *after* their invention and not before.[15] Henry Petrowski argues that it is irritation or dissatisfaction with a design, and not necessity, that spurs most innovation.[16] It would seem that many humans system emerge because they can, and not strictly because they are needed. However, this is hardly a rule.

Pelto does not discuss the origins of the Skolt reindeer herding society, so one cannot trace its emergence. However, Pelto does discuss the emergence of two systems, that of the reindeer herding associations and of the snowmobile. In 1898, legislation divided the reindeer-herding areas of Finland into 57 districts of cooperating reindeer owners, called herding associations. This forced the owners to work together as a system to carry out what were once more individualistic herding operations. Pelto does not mention why these associations

emerged. The snowmobile, on the other hand, emerged from a perceived need. The lack of quick and reliable winter transportation cost inventor Joseph-Armand Bombardier the life of his young daughter. This apparently drove him to create a workable snow vehicle. This eventually led to the invention of a one-person snowmobile, the Ski-Doo, in 1958.

The above examples demonstrate two important aspects of emergence. First, new systems do not emerge into a vacuum, but into *existing* socio-technical systems. The associations emerged as systems within the larger social system of Finland. The snowmobile was introduced to socio-technical systems in which winter travel involved traversing snow. In both cases, the new systems emerged into existing contexts. Secondly, a system cannot emerge from nothing. It must emerge from already existing parts. Since one builds a system by making components interact, the components must *already* exist. One can certainly conceptualize a system without necessarily knowing what all the parts will be, but if the components do not exist, then they must be constructed *before* one can complete the larger system. The associations could not function without the owners just as the snowmobile could not be built without its lightweight engine and skis.

The net evolutionary effect of emergence is to create a new evolutionary unit *and* a new evolutionary context. As a context, the new system determines new selection criteria for its component parts. These parts may have originally evolved in a different context. As such, the new context may radically change the evolutionary direction of that part. For example, once the internal combustion engine became an integral part of the snowmobile, its further evolution *in that context* was constrained by that relationship. The engine, for instance, could not realistically exceed certain weight and size restrictions for most commercial snowmobiles.

As a unit, the new system must *integrate* itself into the existing hierarchy and satisfy the criteria of *its* context if it is to survive. This is especially true of human systems in which people and systems must trade off their use *of* tools with their use *as* tools. The new system may or may not have its own goals, but other units in its context will certainly ascribe it a role, or function, within the existing hierarchy. That role will become better defined as time passes and people become dependent on the new system. Once the role is initially defined, it is further refined through other evolutionary processes which are discussed below.

Phase II: Refinement

As a system integrates itself into its context, it continues to evolve to both better meet its current criteria and adapt when those criteria change. In meeting the challenge of evolutionary survival, an existing system can evolve in many different ways. The are four main ways in which a system can evolve. These are:

a) modification;

b) migration;

c) displacement; and,

d) consolidation.

a) Modification

A system can evolve by modifying one or more of its existing components. For example, the castration of the some reindeer bulls to use as draught animals changed the social structure of the entire herd.

> During mating season the females are gathered together into harems (sometimes numbering a hundred or more) by the strongest and most active bulls. Smaller bulls wander on the peripheries of these harems, occasionally challenging the exclusiveness of the dominant males' sexual control of the females. Meanwhile, the castrated bulls have gathered into groups of their own.[17]

Castrating some bulls, then, changed the herd dynamics.

Modifying one component of a system often necessitates changes to other components in the system as well. Recall that a system is a collection of interacting parts, no *one* of which can be changed.

b) Migration

Often when refining the operation of a system, one adds new components or removes existing ones. This thesis refers to this process as migration, to include both the immigration and emigration of system components. In motorized reindeer herding operations, for example, the herders added nylon ropes to the structure of the permanent roundup sites. In another example, owners of snowmobiles had to acquire tools with which to do emergency repairs. Both of these are examples of immigration. Components sometimes emigrate as well. For example, many Skolts that were once herders turned to wage labour instead. Some of this may have been a result of being forced out of herding, but a few may also have elected to seek alternative employment because it was more lucrative. Certainly some people change jobs in Canadian society not because they were fired but because they saw better opportunities elsewhere. When a system component emigrates to another system, the components of the original system must compensate for the loss.

c) Displacement

Displacement happens when a new component *displaces* an existing component within a system. In Skolt society, the snowmobile displaced the reindeer-drawn sled. Displacement often results from competition among components for a place within the system. To successfully compete, the new component must usually still perform the basic *role* of the original component, but offer some additional advantage. In this case, both the snowmobile and the sled offered means of long-distance travel but the snowmobile was significantly faster, albeit more expensive.

Once again, this can change other aspects of the system as well. The role of skis was not entirely displaced, but significantly changed. While the Skolts seemed to prefer snowmobiles, they often carried skis on long journeys for emergency transportation. Also, the purchase of snowmobiles required the herders to find money for gasoline and repairs, where before they were more self-sufficient.

d) Consolidation

Consolidation is similar to emergence in that it results from the combination of existing units into a new system. This creates an *intermediate* level between those units and the larger context, much like emergence. However, the goal of emergence it to create a new system that can do more as a group than individually, whereas in consolidation the goal is to simplify an existing system by reducing the number of interactions between components.

When a system first emerges, it may be relatively uncomplicated. But over time, mostly through immigration of new components, the system can become more and more

complicated. When it becomes too complicated for humans to manage with existing tools, then one solution is to group, or consolidate, components together into several intermediate systems. This may actually have been the motivation behind the creation of the reindeer associations, to simplify the management of reindeer herding activities for the Finnish government. If so, then the creation of the associations would more properly be called a consolidation rather than an emergence.

As with emergence, these new intermediate levels in the hierarchy are both units and contexts. As units, they must integrate themselves with the other units at that level. As contexts, they may change the selection criteria of their units. In the case of the associations, the Skolts suddenly found themselves in less control over their herding practices because the association was dominated by other groups. They found themselves being selected against in this new context.

Once a system has consolidated, then each new level consists of less units than before and is therefore simpler than before. This allows each level to then evolve through immigration until further consolidations are required and additional levels are installed. In this way, huge bureaucracies begin to evolve.

Of the four processes, modification is least likely to significantly alter the hierarchical structure of the system. On the other hand, components that migrate into a system can introduce many additional hierarchical levels to that system; those that migrate out may eliminate some. Likewise, a component that displaces another can be significantly more or less complex in a hierarchical sense. Lastly, consolidation changes the hierarchical structure of a

system by definition. Despite the many avenues for evolutionary change, however, most if not all systems are eventually selected against.

Phase Three: Extinction

All systems eventually become extinct for two major reasons: displacement and changing context. Sometimes a new system competes with an existing system for a place in its context. If successful, the displaced system can become extinct if it does not find a place in another context. This was the case with reindeer-drawn sleds. The snowmobile displaced them entirely. "By 1967 the use of sled reindeer became a curious anachronism in most areas of northwestern Finland."[18] But systems can become extinct without displacement if processes in the context in which they exist no longer select for them. For example, no one replaced the lead herder in the role of luring reindeer into roundups. The role was simply not required anymore since herders chased rather than lead the reindeer with snowmobiles.

Note however that while a system may become extinct, this does not guarantee that its components will as well. Once again, in a system the parts are greater than a fraction of the whole. If the whole collapses, some of the parts may disperse and join or regroup into other systems. For example, while the job of the lead herder became extinct, the Skolt herder himself did not. He moved on to other activities. In another example, when the Skolts gained their own independent association, the older association became extinct. However, the owners then worked in one of the two new associations.

Since evolution can happen at any level in hierarchy, the three phases described above

apply *concurrently* to all levels in the hierarchy. This complicates the analysis of nested hierarchical systems evolution. However, a more interesting complication results from the dual nature of elements in nested hierarchy.

4.3.2 Implications of the Unit/Context Dichotomy

The unit/context dichotomy of evolving nested hierarchical systems has some interesting implications. Because these systems are nested, changes at one level can affect other levels as well. This thesis refers to this process as the ripple effect. Combined with the interdependence of system components, the ripple effect leads to convergent systems design.

The Ripple Effect

In a nested hierarchy, a change at one level can eventually affect the entire hierarchy because of the unit/context dichotomy of hierarchical elements. Each level, or context, is made up of its evolutionary units. Therefore, if one unit changes, the context itself changes. Since the context changes, other units in that context must compensate as well. Therefore, a change in one unit affects not only its context but the other units in that context. Change ripples *across* that level in the hierarchy .

Now consider that each unit is also a context for the units *it* contains. If a unit changes, then the selection criteria it sets for its units may also change. Therefore, those smaller units may need to adapt to the new criteria. Since they are themselves evolutionary contexts for even smaller units, the process continues indefinitely. Change ripples *down* the hierarchy.

Consider also that the original context is also a unit in an even larger context. If it

changes as a unit in the larger context, then the larger context and the other units in *that* context may change as well. In this way, change ripples *up* the hierarchy.

In the Lapp example, the introduction of the snowmobile eventually resulted in change to the skills of the herders, the structure of the herding teams, and the distribution of wealth of Skolt Lapp society. Change does not happen in isolation. Evolution in any one part of the nested hierarchy can have repercussions throughout the hierarchy. For something new to work in the long run, then, the changes that result must be selected for at *multiple* levels in the hierarchy. The entire hierarchy must eventually adjust to the change or the change will eventually be unsuccessful.

Also, since evolution is happening concurrently throughout the hierarchy, each separate change acts as an epicentre sending waves or ripples of change throughout the hierarchy. Each wave may attenuate and fade away, or amplify and cause massive change. Sometimes these ripples of change are mutually compatible, but often they work at cross-purposes.

Human Interdependence

It is within this criss-cross of hierarchical waves of change that humans must evolve strategies for survival. As entities with their own goals acting as parts of a larger system, units must evolve strategies which reconcile their use *of* tools with their use *as* tools. Arthur Koestler refers to this as an interplay of self-assertive and integrative tendencies.[19] In human systems, this is reflected in two economic concepts: the division of labour and the benefits of trade.

In the division of labour, a given task once performed by one person is subdivided into

multiple tasks each of which a single person can perform. This allows the labour of one individual to be divided among many people instead. The group of people now performing more specialized tasks can hopefully produce more per person as a team than they could as individuals doing everything for themselves. In other words, they work together as a system. For example, in the time it would take for one Lapp to herd a dozen reindeer, a team of herders working together can herd hundreds, if not thousands.

In the benefits of trade, the concept is very similar. Once again, individuals specialize at specific tasks. In this case, however, people trade their output with one another as they seek to acquire all the goods and services they need or desire. For example, Skolt herders and Norwegian merchants would exchange meat for supplies. Each person hopes that specialization increases their per capita production since they can focus all their time on tasks at which they excel. By then participating in trade networks, they hope to exchange their excess production with that of others. If the trade network consists of enough specialized individuals, then as a system it can provide all the desired goods and services for each individual. And if specialization increases each person's per capita output, then they can hopefully acquire more total goods and services than they could if they had to produce everything for themselves. Once again, they are all interdependent parts in a system.

The division of labour and the benefits of trade seem to describe the same process, that of creation of specialized units operating together as a system.. There is a subtle difference, however. In the division of labour, people pool their specialized efforts to produce a single product and then share in the proceeds. In the benefits of trade, people exchange specialized goods and services. This does not, however, preclude the application of the division of labour

to the benefits of trade. This results in the formation of companies which trade the collective work of their employees with one another.

Both the division of labour and the benefits of trade lead to specialization. There is an obvious analogy between specialization in human socio-technical systems and speciation in ecological systems. Whether or not that analogy is appropriate or provides useful insights is left for future research. However, this specialization does have some interesting implications.

If people specialize, then to survive they must participate in a larger network of trade. Specialists, then, do not so much as compete with people with different specialties, but rather cooperate as a system. However, since being part of such a co-operative trade network has advantages specialists *within* a given field may compete with each for a place in the trade network if the system cannot support them all. In this way, herders from different associations may compete with one another even though they all buy supplies from the same Norwegians. However, they may also compete in sense of bartering, of setting prices for their products and services. In selecting a specialty, a person demonstrates integrative tendencies as they prepare to be used as tools by others. In setting prices and competing with others in their field, people demonstrates self-assertive tendencies. One way to describe humans, then, is as "opportunistic seekers of co-operation."[20]

This specialization has a down side as well, however. If the system in which someone specializes no longer has use for that specialty, the specialist will be hard pressed to exchange his or her services for those of others. No herding team needs a bell-reindeer anymore.

The interdependence created by the division of labour and the benefits of trade

combines with the ripple effect to result in a very interesting evolutionary concept: convergent design.

Convergent Design

With a human population of more than five billion people distributed into over 180 countries, why are there so few methods of government in use? Why do most North American houses have a kitchen, a living room, a front foyer, one or more bathrooms and several bedrooms? The answer lies in what James Utterback, professor of management and engineering at MIT, calls dominant design.[21]

When a new, innovative product such as the typewriter or snowmobile emerges onto the marketplace, often no one knows precisely where and how it fits into the existing scheme of things.

> At this stage in a product's evolution, both producers and customers are experimenting. Even as new companies enter with uniquely designed products, established firms are busy perfecting their original designs and introducing new models; and customers are not yet so wedded to any particular design or company that they will not experiment with something new. Industry standards at this stage are usually rudimentary, if they exist at all.
>
> Within this mixture of experimentation and competition some center of gravity eventually forms in the shape of a *dominant product design*.[22]

At this point, firms either converge to produce small variations of the dominant design or quickly go bankrupt. Examples are plentiful. Whether shopping for a manual typewriter or a personal computer, the similarities between brands seem to far outnumber the differences. Each product has its dominant design.

Customer expectations initially determine the dominant design. Once customers figure

out exactly how and where to use a new product, features that were once optional become required. Utterback states that these features become *implicit* to the product design. For example, few people would by a computer without a hardrive, a car without a radio or a house without a kitchen. Any company producing undesired products becomes extinct.

However, once customer expectations determine the dominant design, production economics maintain it. Manufacturers of the product optimize their production facilities to produce that design. Once the production techniques are optimized to produce that particular design, it becomes difficult to radically change either the product design or the production techniques. They are so intertwined as to be inseparable. Changing one means changing the other. Only small, incremental changes to either are possible without greatly disrupting the other. While auto manufacturers may try to stay as flexible as possible, they must still shut down their plants to retool for new models of automobiles, even though the new models are only variations of the dominant design.

A dominant design remains dominant until a major new innovation comes along, such as the electric typewriter. Producers then settle on a new dominant design which incorporates the innovation. Even then, much of the original design remains. Most computers still use the QWERTY keyboard.

Eventually, a radically new innovation will displace the dominant design entirely, as is the case with the snowmobile replacing the reindeer-drawn sled. Even in this case, however, the design maintained many of the previous components. Both the sleds and the snowmobiles have skis, seats and a source of motive power. However, no one would confuse sled with a snowmobile.

While Utterback focuses on technological innovations, his concept can easily be extended to emerging evolutionary units in general, and socio-technical ones in particular. Pick any television channel and watch their newscast. It will likely begin with an overview of the major news events, sometimes giving a brief look at sports and weather as well. The headline news follows. Next on the list is sports. Eventually, the weather person provides a forecast. The networks have all settled on a fairly consistent dominant design. When one channel succeeds with something innovative, other channels quickly adopt it into their own broadcasts. Many news casts now include an entertainment segment as well. The similarities between newscasts are usually more numerous than the differences.

Dominant design is a side effect of the ripple effect of hierarchical evolution and the interdependence of system components . When a new unit emerges, it emerges into an already *existing* hierarchy. Every newly emerged evolutionary unit must integrate itself into the that hierarchy if it is to survive, adapting to the selection criteria determined by its larger context. As a new unit integrates itself into the hierarchy, it starts to perform a *role* in that hierarchy. Other units in the hierarchy begin to *expect* certain things from the new unit. As the new and existing units become more and more interdependent, their mutual expectations force the new unit to converge towards a more or less stable design. This may, in turn, necessitate the existing units modifying their own design as well. This *convergent* design is the more general case of Utterback's dominant design. Convergence can be viewed as a refinement of the original emergent unit as it integrates itself, or is integrated, into the existing hierarchy.

Once fully integrated, further major changes to the convergent design are risky and difficult to achieve because of the ripple effect. For the modified design to succeed, the

selection criteria of its context must select for and not against the changes. Even if it succeeds in its immediate context, the changes that ripple up, down and across the hierarchy as a result of the change may eventually be selected against at another level.

To further complicate the evolution of these convergent designs, most exist simultaneously in more than one nested hierarchy. Each hierarchy generates its own selection criteria which often clash with those of the other hierarchies. Consider the snowmobile. To how many hierarchies does it belong? First, it is part of a corporate hierarchy as a product. Within the corporate hierarchy, there are marketing, engineering, accounting and production hierarchies, to name but a few. The manufacturers as a whole must operate within a larger economic context, with customer demands and government regulations. The snowmobiles must also function within the existing support hierarchies. Drivers must have access to fuel and parts for repair. All these hierarchies have their own, often contradictory, selection criteria which *directly* affect the survival of the snowmobile and determine the success or failure of any new changes. Any changes must successfully integrate into all these systems. Is it any wonder that, especially in more complex systems, any newly emerged unit that manages to successfully attain a convergent design will not change significantly until a truly innovative product comes along to displace it?

However, even if it does not change significantly, change it must if it is to survive. Since the convergent design is but one element in a nested hierarchy, its context is still subject to ripples of change from other sources. If the convergent design does not adapt to these changes, it may face extinction even without competition from other designs. This is the paradox of hierarchical evolution. Change is both risky and necessary for long-term survival.

A Brief Note on Divergence

The above discusses focuses on the convergence in design. There are, of course, designs that diverge as well. This is primarily the result of a change in context. As a unit moves from one context to another, the selection criteria it must meet likely change as well. This will eventually result in a design that diverges from the original but converges to new dominant form. For example, designing cars for military use resulted in the jeep.

4.4 Summary

This chapter examined patterns of nested hierarchical evolution of human socio-technical systems using Skolt Lapp society as a case example. It identified three distinct phases of systems evolution: emergence, refinement and extinction. Refinement can take one of four forms: modification, migration, displacement or consolidation. These process of change operate at all levels in the hierarchy concurrently. As a result of the unit/context dichotomy of hierarchial elements in an evolving system, changes at one level can ripple up, down and across the hierarchy. Combined with the interdependence of system components, this results in convergent design.

The next chapter discusses some of the implications and insights of this model of evolutionary change. It also presents some suggestions for future work in the model's development.

1. Pertti J. Pelto, The Snowmobile Revolution: Technology and Social Change in the Arctic (Menlo Park, California: Cummings Publishing Company, 1973), iii.

2. Pelto, 11.

3. Pelto, 21.

4. Pelto, 23.

5. At the time of Pelto's book, all herders discussed were male.

6. Pelto, 5.

7. Pelto, 37.

8. Pelto, 41.

9. Pelto, 45.

10. Pelto, 97.

11. Pelto, 101.

12. Pelto, 91.

13. Pelto, 99-100.

14. Pelto, 98.

15. George Basalla, The Evolution of Technology (New York: Cambridge University Press, 1988).

16. Henry Petrowski, The Evolution of Useful Things (New York: Vintage Books, 1992).

17. Pelto, 36.

18. Pelto, 74.

19. Arthur Koestler, The Ghost in the Machine (Hutchinson & Co, 1967; reprint, London: Arkana Books, 1989), 56.

20. Staff writer. "Evo-economics," The Economist (December 25th 1993): 95.

21. James M. Utterback, Mastering the Dynamics of Innovation (Boston: Harvard Business School Press, 1994).

22. Ibid., 24.

Chapter Five

Discussion

5.0 Introduction

Chapters One through Four of this thesis present a modelling framework with which to analyze change in human socio-technical systems, that of the evolving nested hierarchical system. The model helps one understand change in these systems, but it would be useful if the model also suggested methods for directing that change in a positive way as humanity strives for continued survival. This chapter looks a bit closer at the implications hierarchical structure has for evolutionary survival and suggests two approaches to aid in changing the evolutionary futures of human systems. It then discusses the need for future work to better develop the model of nested hierarchical systems evolution.

5.1 Isolation and Survival in Evolving Nested Hierarchies

Chapter Three argues that the consolidation of system components into new, intermediate levels simplifies the entire hierarchical system by reducing the total number of connection possibilities between components in the hierarchy. As the number of components in human systems grows, more and more levels emerge as humans attempt to retain some sort of control over the system. This leads inevitably to large bureaucracies with many levels of management. These additional levels separate low level units farther and farther from high

level contexts. Recalling Kelly's distinction between clockwork and swarm systems, simplifying systems to retain some level of control comes at the cost of a rigid and inflexible organization. There is a tradeoff in determining the optimum number of intermediate levels. Regardless of where one finds the balance, however, some intermediate levels will emerge.

Since these intermediate levels are simultaneously both evolutionary units and evolutionary contexts, they change the evolutionary pathway of the system. The evolution of the units that the level now contains was previously directed by selection processes of the larger context. Now their evolution is directed by the context of the intermediate level. It is the level *as a whole* that now evolves within the larger context.

Consider what this implies for the units now *separated* from the larger context by the intermediate level. The units must now evolve strategies for survival in the new intermediate context rather than, as they had before, for survival in the larger context. As additional intermediate levels emerge, the units become more and more isolated from their original context. The strategies they continue to evolve become more and more dependent on their current level in the hierarchy rather than on their original context.

Consider human social evolution as an example. Long ago, most humans lived in small systems known as hunter-gatherer societies. There was typically only one level separating individuals from their local ecosystems, that of their social group. The strategies these people evolved for survival were greatly influenced by their local ecosystems as too much consumption in any one location would permanently eliminate their food supply.

Compare that situation with life in modern Canadian society. Few people in Canada depend directly on their local ecosystem to provide food and other necessities. Instead, they

rely on getting enough money to purchase these necessities. But how do most people earn money? They earn money by being parts of multi-levelled socio-technical systems such as corporations and government agencies. Such organizations are themselves parts of even larger hierarchical systems, such as cities, provinces and countries. Even though, physically, we are still parts of local and global ecosystems, these socio-technical organizations isolate most people from their ecological contexts through many intermediate, socio-technical levels. This results in the evolutionary strategies of individual humans being more economically than ecologically driven. While necessary for short-term survival, this can have dire consequences as humanity is now discovering.

To summarize, to continue to survive in a given intermediate context, two conditions must be met.

1) The unit must continue to be selected for in the context of the intermediate level.
2) The intermediate level itself must continue to be selected for in *its* larger context.

If the unit is selected against in the context of the intermediate level, then it must either find a place in a different system or become extinct. If the *intermediate level* as a whole is selected against, then *it* must find a place in a new system or become extinct. If the level, as a unit, becomes extinct, then the units it contains must move on or face extinction as well. For example, a welder that is laid off from one firm might find work in another or rely instead on government financial assistance. If the construction firm loses its customers, it must find others or cease operation. If the firm for which the welder works fails, the welder must again

find income elsewhere.

As the number of intermediate levels in the hierarchy increases, the lower level units become dependent of the survival of more and more systems. If any one of these systems fails, all lower level systems must adapt or face extinction as well. As long as even higher level contexts exist, however, there is always the opportunity to adapt and move on. Unfortunately, many of the strategies humans have evolved to survive at lower levels in their socio-technical systems are incompatible with the long-term survival of the larger socio-technical organizations. Most notably, many individuals and organizations are concerned with the current environmental crisis. However, as long as the strategies needed to survive at lower levels in the hierarchy continue to be inconsistent with long term survival of the larger organizations in the global ecological context, this crisis will continue. If the situation does not change, then eventually the global ecosystem may select against human society *as a whole.* The implications for humans as individuals are unknown, but likely undesirable.

5.2 Changing the Situation

Humans have an advantage over most other species in that they can, if they choose to do so, look beyond their immediate level in their hierarchy. Although they must still evolve strategies to survive in their immediate, socio-technical context, this does not preclude the possibilities of developing strategies that are also compatible with the long-term survival of the larger organizations. Sometimes, however, people must be motivated by others to change their behaviours for the good of society as a whole.

There are two main ways one can influence the evolutionary strategies of individual

units. The first involves changing the context in which the units operate, thereby changing the selection criteria which guide the evolution of strategy. The second approach involves getting the units to change their strategy within the existing by showing them that the new strategy is actually an improvement. A combination of both approaches is also possible.

Changing the Context

Changing a unit's evolutionary context changes the selection criteria by which selection process judge that unit. If evolutionary units, such as humans and their tools, really do evolve strategies to survive in their given evolutionary context, then changing the context may force them to evolve new strategies to survive in the new context. For example, to encourage manufacturing firms to develop environmentally sounder practices, the government could create legislation penalizing companies with poor environmental practices and offer grants to companies with good environmental practices. Hopefully, this change in context would select for environmentally friendlier firms by making cleaner practices more profitable.

This approach has the advantage of letting the firms themselves do most of the work to change strategies. However, there are drawbacks as well. First, it may be difficult to actually change a context in a way which encourages the desired changes, especially if the context is large compared to the agent of change. Enforcing legislation and financing grants on a national level are both expensive ventures. The government may not have enough resources to significantly alter the context, in which case the firms have no reason to change. Second, even if the government has enough resources, those firms that do not change may be selected against and become extinct. While this achieves the goal of eliminating environmentally unfriendly

companies, the corresponding loss in employment may be very undesirable.

If one wishes to change a context, then one requires a good understanding of the existing context, its associated selection criteria and its associated selection processes. In the case of humans and their tools, human beings typically act as selection processes. It is important to remember when dealing with humans as agents of selection that they operate on the basis of *perceived* selection criteria. In Finland, there was great debate in the 1960s as to whether or not snowmobiles were indeed effective for reindeer herding. These perceived selection criteria may or may not reflect the context's actual selection criteria. This is determined in the long run when those units whose strategies do not address the real criteria do not survive. Since units are often responding to perceived rather than actual criteria, sometimes it is not necessary to change the context if it should already engender the desired strategies. Instead one might be able to convince the units that they are operating under mistaken perceptions of the real criteria.

Changing the Strategy

An alternative approach is to leave the context as is and demonstrate to the units that a different strategy actually helps them better meet the selection criteria of their context and at the same time meet higher level criteria as well. For example, the government might convince firms that environmentally friendlier practices will lead to both increased sales due to consumer support and reduced costs due to less resource consumption. This approach avoids the difficultly of changing the context but may involve more effort at the level of the individual firm. It also has the drawback that it does not guarantee that the firm will maintain these

practices if an even newer and less environmentally friendly approach proves to enhance their survival even more.

If one wishes to suggest alternative strategies, however, one needs to understand not only the current strategies of units in question, but also the *perceived* selection criteria for which these strategies emerged.

Since neither approach by itself may achieve the desired result, it is often prudent to combine approaches. For example, the government might change the context by increasing public awareness of the environment and then demonstrate to firms that implementing environmentally friendlier practices is to their competitive advantage. Both approaches, however, require a better understanding of nested hierarchical systems evolution than the thesis presents in the previous four chapters.

5.3 Future Work

As stated in its introduction, the goal of this thesis is to lay the foundation for a different perspective on change. Laying a foundation is but the first stage to building the overall structure of a model. There is much work still needed. Three main areas for future work are the analysis of more in-depth cases, the inclusion of complex systems theory and the incorporation of the concept of feedback.

5.3.1 Case Studies and Partitioning

Chapter Four describes a brief example of how to apply the model to an actual case. Much more detailed case studies are required to further test and refine the model of an evolving nested hierarchical system. The last chapter also introduces some very basic partitioning guidelines based on the concept of middle-out modelling. In performing more detailed case studies, it would be helpful to have correspondingly more defined partitioning guidelines.

5.3.2 Complex Systems Theory

While Chapter One discusses the concept of system in some detail, it leaves untouched the ideas of complex systems theory such as non-equilibrium thermodynamics, catastrophe and chaos theory, and feedback. These ideas have important implications for evolutionary models. While it is beyond the scope of this thesis to pursue these ideas in detail, they do raise several issues that should be pursued in future work.

Non-equilibrium Thermodynamics

While this thesis used mostly logical arguments to support the possible existence of hierarchies, there are thermodynamic arguments as well. The following discussion assumes some knowledge of thermodynamics. Readers who wish additional information are asked to refer to the referenced material.

In thermodynamics, systems seek equilibrium, a state in which there are no gradients. Moving a system away from equilibrium requires the application of gradients, such as the massive temperature difference between the earth and the sun.

> The thermodynamic principle which governs the behaviour of systems is that, as they are moved away from equilibrium, they will utilize all avenues available to counter the applied gradients. As the applied gradients increase, so does the system's ability to oppose further movement from equilibrium.[1]

Bénard cells are an example of this process of gradient opposition.

This property of systems can lead to hierarchical structure in two related ways: dissipative structures and imposed constraints. Dissipative structures are structures that emerge to dissipate gradients and return the system to equilibrium. These include structures such as tornadoes and whirlpools. Convection currents, such as those found in Bénard cells, are another example. Schneider and Kay "suggest that life exists on earth as another means of dissipating the solar induced gradient."[2]

Life dissipates the solar gradient by degrading the *exergy* content of the incoming solar *energy*. In simple terms, exergy is a measure of how much work one can get out of a given amount of energy. Even though thermodynamics tells us that energy is conserved, not all energy is equal. For example, one can use a fire to keep warm by tapping the energy stored in the wood. However, once the wood is burned, their is no more warmth to be extracted. The energy was not destroyed, merely transformed. However, the usefulness of the energy, its exergy, is destroyed. Thermodynamics also tells us that exergy degradation is an irreversible process.

Interestingly enough, the living dissipative structures which emerge to dissipate the solar gradient are themselves non-equilibrium systems. As they dissipate the solar energy through exergy destruction, they also store exergy such as the exergy contained in fire wood. In dissipating the larger solar gradient, they create in themselves smaller, localized gradients.

One implication of this process is the emergence of food webs as the system attempts to eliminate these gradients as well. Another implication, more relevant to this thesis, is the possibilities for hierarchies.

This living non-equilibrium systems can exist because of thermodynamic constraints imposed by their structure. Hatsopolous, Keenan and Kestin propose that

> when an isolated system performs a process after the removal of a series of internal constraints, it will reach a unique state of equilibrium: this state of equilibrium if independent of the order in which the constraints are removed.[3]

This statement has many implications, one of which is that if a system is not at or moving towards its unique state of thermodynamic equilibrium then there must be internal constraints prevented it from doing so. These constraints allow for a constrained equilibrium. This thesis suggests that nested hierarchical structures can emerge using these constrained systems as building blocks.

External gradients are not necessary to maintain the internal gradients. Wood stores exergy as a living tree and as chopped firewood. However, external gradients are necessary to store the internal gradient in the first place. Also, there is a definite qualitative difference between a living tree and a dead log. Life, it seems, needs external gradients to survive. Exergy is just one, albeit perhaps the most important, of these gradients. Biology requires water and other nutrients as well. Human society requires additional raw materials. Systems can recycle nutrients, but this comes at the cost of exergy. Exergy cannot be recycled. This has important implications for the sustainability of human society.

This section touches only briefly on non-equilibrium thermodynamics, but hopefully it demonstrates the many possibilities for future work in this area for improving our

understanding of evolving nested hierarchical systems.

Catastrophe Theory and Chaos Theory

These theories involve too much to explain quickly here. Incorporating them into the modelling framework of this thesis will require much additional work. It is important to do so, however, because these theories tells us that uncertainty is unavoidable, that we cannot in principle predict the future. However, they do allow for some insight into the possibilities for the future as well as some qualitative descriptions of system behaviour. What implications these theories have for evolving nested hierarchical systems is left for later work.

Feedback

Feedback is an important system concept and needs to be more fully incorporated into this thesis in future work. Feedback is a cyclical process. There are two types of feedback: positive and negative. Positive feedback describes self-reinforcing processes. Investing money in a compound interest account is a good example. If left untouched, the principle invested will grow exponentially as interest feeds more money into the account. Negative feedback describes self-weakening processes. An annuity with a set term is a good example. As each payment is made, there is less money in the account with which to earn interest and the balance slowly dwindles. Eventually, all the money is paid out.

When negative and positive feedback cycles are combined, *homeostatic* systems can emerge which maintain a given state. For example, a permanent annuity can combine the positive feedback of interest with the negative feedback of withdrawals to maintain a set

balance. *Autopoetic* systems extend homeostatic processes to provide the ability to repair damage, among other things. Autopoetic systems typically involve living components. To continue with the above examples, one could imagine an annuity with an administrator that adds money to the fund to compensate for low interest and removes money to compensate for high interest, thereby maintaining the set balance.

One can view selection as a feedback process and, as such, the ideas introduced here need to be incorporated in the model of a nested hierarchical evolving system.

5.4 Summary

Increasing numbers of intermediate levels isolate low level evolutionary units farther and farther away from their original context. The survival strategies of this units become more and more dependent on their current context and less and less dependent on their original context. However, the original context still selects for and against the intermediate levels which contain these units. If the strategies of the units result in higher level systems which are incompatible with the original context, then the entire hierarchy can be selected against. One sees evidence of this in North American society where individuals are more concerned with surviving economically than ecologically.

One can encourage change in the strategies and behaviours of individual units in two ways. The first changes the units' context and associated selection criteria in the hopes that he units will evolve more desirable strategies before becoming extinct. Alternatively, one could present different strategies to the units that compatible with both low level and high level selection criteria. One could also combine these methods.

Both approaches require a better understanding of nested hierarchical systems evolution than the thesis presents. Areas for future work include the analysis of more in-depth case studies and the incorporation of complex systems ideas such as non-equilibrium thermodynamics, catastrophe and chaos theory, and feedback. Thermodynamics tells us that real processes are irreversible. Catastrophe and chaos theory tell us that everything is uncertain. Feedback tells us that process can self-reinforce or self-weaken, and can combine to do both concurrently. These ideas need to be further explore in the context of nested hierarchical systems evolution because they imply that all approaches we take to direct change are made under conditions of irreversibility and uncertainty, and can feedback on themselves.

1. Eric Schneider and James Kay, "Life as a Manifestation of the Second Law of Thermodynamics," pre-print from Advances in Mathematics and Computers in Medicine (1994): 6.

2. Schneider and Kay, 15.

3. Schneider and Kay, 5.