A Fundamental Systems Hypothesis relating Resources, Risk, Complexity and Expected Output Value

James Bradley
Department of Computer Science
University of Calgary,
Calgary, Alberta, Canada

Abstract A working hypothesis is presented and justified, called the Fundamental Systems Hypothesis. It relates expected net output value, complexity, risk and resources, and governs all human-agent-directed systems. The general veracity of this Hypothesis appears such that it could be considered a Fundamental Law of Systems. The risk measure can be either conventional standard deviation risk or mean deviation risk. There are two risk parameters: positive and negative risk. There are two complexity parameters: monitoring or checking complexity, and resource-sharing complexity. Monitoring complexity is defined as a specification length, and measures complexity in the system's environment-coping procedure that monitors a time function representing the unfolding environment. Resource-sharing complexity measures the execution time of a complex resource-sharing procedure. The Hypothesis is expressed as a mathematical relationship that reduces to numerical values for specific system circumstances. It also quantifies real economic losses, and gains, associated with system risk. The established Markowitz-Sharpe-Lintner relationship between return, capital resources and risk for the subclass of financial systems is inherent in the Hypothesis. The Hypothesis can be subjected to experimental test.

Key words Complexity, coping procedure, fractal, Hurst constant, time function, random walk, resources, risk.

Introduction

In recent decades the complexity of systems, particularly computer systems, has very greatly increased. As well, complex systems are increasingly having to cope with risk in their operating environment. Examples are systems in aircraft and ships, both military and commercial, systems in spacecraft and orbital vehicles, systems in power plants, especially nuclear plants, systems in hospitals, systems for general surveillance, and systems in critical manufacturing processes. All of these systems are ultimately human-agent directed systems. Unfortunately, these agents may not always be able to reason clearly about the relationship between system output, resources, environment risk, and complexity. As a result serious mistakes can occur, both in system design and system operation. The purpose of this paper is to report the results of research that clearly reveals the nature of this relationship, in order to help prevent such mistakes.

In this paper we argue for a working hypothesis, called the Fundamental Systems Hypothesis, relating system output value, complexity, risk and resources, and governing all human-agent-directed systems. However, we are confident of the general veracity of this hypothesis, so that it could also be considered as a Law of Systems. Persons familiar with its workings will be better equipped to think about, analyse, create and manage, computer systems of all kinds.

There seems to be little in the literature on computer systems, of work done to relate complexity and risk. However, there is a very great deal in the financial literature on research relating financial returns to risk and financial resources. The ground-breaking work in this area was done by Markowitz [18], Sharpe [23, 24] and others [7, 13, 14, 27] in the 1950s and 1960s, and their results are widely used in the financial arena [3, 19, 21]. The work described in this paper first of all extends Markowitz's and Sharpe's work to how system output in general is related to risk and resources.

1.0 A Fundamental Systems Hypothesis

We consider a system to be an entity that functions under the direction of one or more agents, and which employs resources R, often portable, in an environment E of unalterable charcteristics, to convert a set of inputs (N), informational or physical or both, to a set of outputs (U). The system outputs are normally of greater value to the agents than the system inputs, so that is there is a net positive value (V) to the agent, and measurable by the agent, associated with system operation.

To begin, we state the Hypothesis as a relationship between net output value to the agent, resources, environment risk and system operational complexity as follows:

$$V = K\,R(1 + b_1 r_P - b_2 r_N + b_3 P + M(c_m) + S(t_s)) \qquad (1)$$

which under certain conditions relating to complexity, reduces to:

$$V = K\,R\,(1 + b_1 r_P - b_2 r_N + b_3 P + a_1 c_m + a_2 t_s) \qquad (1a)$$

* V is the dependent variable, with units of value, equal to the expected net value per unit time produced by the system.
* R is an independent variable, with units of value, equal to the value of the resources (often portable) of the system directed towards exploiting a physical environment.
* r_P is an independent variable, with no units, equal to a measure of the net risk inherent in the system's operating environment of loss of value, for risk it can pay to run repeatedly (positive risk); the independence of r_P is such that it can be varied by the agent shifting the system (resources) to a different operating environment; only either net positive or net negative risk (see r_N) can be present in an environment at any given time, and not both.
* r_N is an independent variable, with no units, equal to a measure of the risk inherent in the system's operating environment of loss of value, for net risk it can not pay to run (negative risk); $b_2 r_N$ will frequently be zero; the independence of r_N is such that it can often be varied by the agent shifting the system to a different operating environment.
* P is an independent variable equal to the value of resources directed exclusively to risk prevention (or reduction) as experienced by the system. Total resources in the system are thus R + P. P is usually zero, and included in R, and non-zero only where explicit

demonstration of the effect of risk reducing resources P in a specific environemt is required.

*c_m is an independent complexity variable, the monitoring complexity, with units of length; it is equal to a measure of monitoring or checking complexity, in one or more environment-coping procedures that are part of the system. $M(c_m)$ is a function, the *monitoring complexity schedule*, whose value is zero when $c_m = 0$; it increases with c_m at a decreasing rate until $M(c_m)$ reaches a saturation value, where further increases in c_m have no effect. Far from saturation, it can be convenient to assume that $M(c_m)$ can be approximated by the linear function a_1c_m, where a_1 is a constant, as in (1a).

*t_s is an independent complexity variable, the complex resource-sharing procedure time, with units of time. $S(t_s)$ is a function, the *resource-sharing complexity schedule*, whose value is zero when $t_s = 0$; it increases with t_s at a decreasing rate until $S(t_s)$ reaches a maximum value and resource utilization is 100%, where further increases in t_s have no effect. Far from this maximum, where t_s is small, it can be convenient to assume that $S(t_s)$ can be approximated by the linear function a_2t_s, where a_2 is a constant, as in (1a).

* a_1, b_1, a_2, b_2, and b_3 are constants, for a specific class of system and environment; as will be explained later, for any environment, the value of either b_1 or b_2 or both must be zero.

* K is normally a positive constant, except in rare cases with isolated systems where supply elasticity is sensitive to net physical/informational output. An expression for K, applicable in such rare circumstances, is given in Appendix 1.

From the Hypothesis expression (1), three important subsidiary Hypotheses can be deduced. If K_3 is an appropriate constant, these are:

(a) $$V = K\,R\,(1 + b_1r_P - b_2r_N + b_3P) \qquad (2)$$

or frequently $$V = K\,R\,(1 + b_1r_P) \qquad (2a)$$

for the relationship between output value, resources and risk, at constant complexity;

(b) $$V = K\,R\,(1 + M(c_m) + S(t_s)) \qquad (3)$$

for the relationship between output value, resources and complexity, with constant risk;

(c) $$V = K_3\,(1 + b_1r_P - b_2r_N + b_3P + M(c_m) + S(t_s)) \qquad (4)$$

for the relationship between output value, risk and complexity, with constant resources.

We look at these three in turn below and argue for their general veracity as the basis for the veracity of (1).

Measure of resources and output value, and resource definition

The exact meaning of the resource measure R (and P) and net output value V is important. The units of both R and V are the same, namely units of value to a human agent, that is, units of economic value. A value is that which is conducive to the survival and well-being of a living organism, for example water is a value to a plant, or shelter is a value to a human. This paper concerns values to humans. Humans, as agents, create and operate systems to help them achieve values. The quantity of any commodity, considered as a standard, that freely exchanges for a unit of (material/informational) output, can serve as a measure of value of any unit of output, that is, its price.

A resource R (or P), to a human agent, may be defined as the result of a set of operations or process, carried out in the past by humans, or by another natural agency, or by a combination of both, where this result either makes possible or enhances repeated execution of current operations that are directly or indirectly conducive to the survival and well-being of humans; futhermore this repeated execution of current operations can be for a limited time, in the case of either a depleting (e.g. mineral deposit) or depreciating resource (e.g. a disk drive), or for an indefinite time, in the case of a sustainable resource (solar energy).

Because V in expression (1) denotes value of output and not units of physical/informational output, there is an economic dimension to the expression which, although of secondary importance from a systems viewpoint, cannot be ignored. Suppose risk and complexity (to be defined later) are zero for a system with resouces R, then (1) reduces to V = KR. To take a self-evident example, according to this expression, if we double the number of apple trees and associated land resources (R), we double the value (V) of the physical (apples) output. But, the economist will object, although you will double the quantity of apples output (U), V will double only if the price per apple is unchanged, for the operation of the Law of Supply and Demand may well result in the increased supply overwhelming demand, causing the price per apple fall, so that V might well falls as R rises, contrary to (1). The economist is right. But normally, continuing with this obvious example, for a single apple tree collection in a large economy, doubling the trees and thus doubling the quantity of apples output will not effect the price per apple, so that V = KR will hold. But clearly, for a given system, K is a constant only where a change in the resources R deployed does not affect the price per unit output by the system (due in technical economic terms to a change in the supply elasticity), which will be the case for systems used in any large economy. [An expression relating K to supply elasticity, relevant in the case of a very small economy or single isolated system, is in Appendix 1.]

2.0 Risk versus resources

Risk of loss of anything of value is normally run to secure some gain in a value, and is therefore of primary interest to humans. Consequently, the concept of risk has been thoroughly studied in the insurance and financial industry [2, 5, 7]. A major result is that we know that there are only two basic methods of dealing with risk to secure the associated gain: either run the risk or do not run the risk. These two methods lead to a deeper classification as follows:

A. Running the risk

(a) Run the risk directly without insurance, and accept the consequences, that is, suffer losses of value when the hazard occurs.

(b) Run the risk directly but insure it too, in whole or in part, and so distribute over time (at least some of) the losses due to the hazard occurring. Some risks cannot be insured.

(c) Run the risk directly but counterbalance it by operating in two areas that behave oppositely, so that what is lost in value in one is counterbalanced, at least partly, by a gain in a value in the other. Some risks cannot be counterbalanced.

B. Not running the risk, while retaining the gain possibility

(a) Avoid the risk, by finding an alternative activity that still achieves the gross benefits (even at some additional cost) but eliminates running the risk; for some risks there is no such alternative.

(b) Prevent the risk, by applying resources

1. From detailed knowledge of the nature of the hazard, apply relevant material resources to eliminate the threat of the hazard occurring. For some risks this is not possible. [This is the principle behind deadlock prevention in computer operating systems. In lay terms, "build the ship so that it cannot sink".]

2. From detailed knowledge of the nature of the hazard apply material resources to prevent the loss should the hazard occur. For some risks this is not possible. [This is the principle behind enhanced deadlock recovery in operating systems. In lay terms, "skimp on building the ship, so that it can sink, but have plenty of lifeboats".]

(c) Neutralize the risk, either by applying a complex environment coping procedure to detect early hazard warnings in time to undertake preventative action, or exploit compensating opportunities. For some risks this is not possible. [This is the principle behind deadlock avoidance with the Banker's Algorithm in operating systems. In lay terms, "skimp on building the ship so that it can sink and has too few lifeboats, but have an effective above/below surface lookout system so that there is always time to take action to avoid a collision with anything".]

Simple risk measures

From classification B(b) above, it is clear that risk has two components, namely the probability p of a hazard occurring and the size L of the loss should the hazard occur. However, in a situation where there is exposure to possible loss (risk), there will be exposure to possible gains in addition. Both rain and sunshine occur. Suppose a situation with loss probabilities p_1, p_2, ... and loss sizes L_1, L_2, ... per unit time period, with a total chance of loss (risk) of therefore $(L_1p_1 + L_2p_2 + ...)$. Suppose in addition gain probabilities q_1, q_2 ... with gains G_1, G_2 ..., so that there occurs an offsetting chance of gain $(q_1G_1 + q_2G_2 + ...)$. In such a situation the risk appears to be less than it would otherwise be if the gains were not possible, for it can be argued that at least some of the losses will have been offset by the gains. However this appearance can be an illusion.

The essence of understanding the nature of risk in general, and developing an adequate measure, lies in the nature of how offsetting gains affect risk. Even if on average

$$(p_1L_1 + p_2L_2 + ...) = (q_1G_1 + q_2G_2 + ...)$$

so that there is no net loss per unit time period, because the gains and losses will not occur at the same time - in one unit of time it may be all losses and no gains and vice versa - it is clear that while the offsetting gains may ameliorate the risk to some extent, they do not eliminate the risk. Instead the observable result of this stream of unpredictable losses and gains will be a fluctuating V with time. The solution to the problem of a risk measure that allows for unpredictable gains as well as losses in any system case corresponds to the risk measure developed in the 3rd quarter of the century by H. Markovitz [18] and W.F Sharpe [23,24] and others [7, 13, 14, 27] for financial systems. This solution is widely accepted and used in practice with financial and corporate systems [3, 5, 7, 21], and is believed to be fundamentally correct. We will develop it further for the case of any system, financial or otherwise.

Deviation Risk Measures

Suppose that over a long period of time T, involving many time units (e.g. years, minutes) the total net value output of a system is W. Suppose that over the same time the system was exposed to unpredictable losses and gains in value, and that the results of all of the realized losses and gains are included in W. Assuming that output at the end of each time unit was never used (reinvested) as part of the system resource R in the next time unit, then the average or expected net output value V per time unit must be W/T.

Although the average output (or expected output) value may be V, the actual output values in each of n time periods during T will be something like:

$$V - L_1, V - L_2, ... V - L_i , V + G_1, V + G_2, ... V + G_j$$

where L_1, L_2, ... are deviations downward (losses) from the average output value V, and G_1, G_2, ... are deviations upward (gains) from V, so that the expected or average output actually rarely occurs. These outputs can occur in some unpredictable order, and we will have

$$j + i = n, \text{ and}$$

$$(L_1 + L_2 + ...L_i) = (G_1 + G_2 + ...G_j)$$

since we must have:

$$(nV + (G_1 + G_2 + ... + G_j - L_1 - L_2 ... -L_i))/n = V$$

Thus in practice with any system we must deal with unpredictable losses (L_1, or L_2, ...) and gains (G_1, or G_2, ...) with respect to an average output value V in a given time unit (year, minute, etc), and it is such losses with respect to an average or expected output value V that must be used in the measure of risk.

In the context of a fluctuating output and an average output per unit of time, for a meaningful measure of risk there are now two main choices:

Choice 1. Take the mean deviation of (L_1, L_2, ... G_1, G_2, ...) to give the Mean Deviation (MD) risk measure with respect to expected output value V.

Interpretation: An MD-risk of m means that over the next time unit (year, month) there is a 50% chance of a loss, with respect to the expected average output V, equal to m, and that to a good approximation, there is a 25% chance of a loss less than m and 25% chance of a loss greater than m.

Choice 2. Take the standard deviation of (L_1, L_2, ... G_1, G_2, ...) to give the Standard Deviation (SD) risk measure with respect to expected output value V.

Interpretation: A SD-risk of s means that over the next time unit, there is a 50% chance of a loss with respect to the expected V, and 34% chance of a loss $< s$ and a 16% chance of a loss $> s$; in addition there is a 47.5% chance of a loss $< 2s$ and a 2.5% chance of a loss $> 2s$. In addition there is a 68% chance of either a loss $<s$ or gain $<s$, with respect to the expected V. The percentages assume that losses and gains in each time unit are distributed reasonably normally.

It is clear either deviation risk measure could be used, as they are closely equivalent. The MD-risk measure has the advantage of being more intuitively appealing when seeking insight into a situation, whereas the SD-risk measure is more mathematically tractable when involved statistical analysis is necessary. The SD-risk measure is widely used in data processing relevant to portfolio selection and management in investment analysis [3, 5, 18, 21, 23] Since this paper is seeking insight into fundamental matters, we will tend to discuss matters with the MD-risk measure in mind; however the same results will be obtained if the reader relies on the SD-risk measure. Note that when risk is expressed as a fraction of R, as is common, it has no units.

The influence of risk on output value

We must now show using MD (or SD) risk measures r_P and r_N, that

$$V = K_1R(1 + b_1r_P + b_2r_N + b_3P)$$

where only one of b_1, b_2 is non zero. A somewhat similar but simpler expression is used in finance and developed by Sharpe [23, 23] to relate the income (I) from an invested principal sum S under conditions of risk, namely

$$I = S(1+ i + br) \qquad ...... (5)$$

where i is the rate (expressed as a fraction of S) of return, or interest rate, in the absence of risk, and br is the extra rate of return added by the presence of risk (SD-risk) r in the environment. This expression can be rewritten

$$I = (1 + i)S(1 + rk/(1+ i)), \quad \text{or} \quad I = KS(1 + Br) \quad (5a)$$

where K and B are constants. This converts expression (5) to an expression in the same form as the Hypothesis-derived expression (2a) or $V = KR(1 + br_P)$ with negative risk r_N assumed zero and risk prevention resources P assumed zero. This indicates that the financier's risk r is the equivalent of the positive risk r_P for systems in general. This relationship (5) is basic for financial systems. In general, the rule that return on investment of a given principal sum S increases with the associated SD-risk of the financial environment selected for investment, or equivalently, that for a given income, less S is needed as environments with increasing risk are selected, seems to hold true in the environments of financial systems [3, 24].

We can now show that there is a reason for a positive linear relation between system output value and the the risk of the operating environment, applicable to all systems, can be deduced from the existence of only two fundamental choices in dealing with a risk associated with a system, namely (A) run the risk, or (B) do not run it, while preserving the associated gain, as earlier classified.

Consider now two environments E_1, and E_2. E_1 is risk free and gives a fixed value output F per time unit for application of a system with resources R, which regularly transforms inputs N to outputs U, either physical, for example a system that transports parcels from city A (input N) to city B (output U), or informational, for example a system that transmits computer files from file server A (input N) to file server B (output U). E_2 is the same as E_1 except that in E_2 the transformation of N to U is in a positive risk environment.

Suppose gross output value in E_2 is F + G per time unit, each time the risk in E_2 is run but it just happens that the hazard does not occur. Thus G is a gain in gross output only if the hazard does not occur when the risk is run. The agent for this system in E_2 can achieve the routine input-output transformation N to U and generation of the gross output value F + G by either (1) running (accept exposure to) the risk or (2) by finding a way of not running it while preserving the transformation N to U, the value output F, and the extra gross value output G. But either way, on average, there will be an unavoidable cost with respect to the extra gross output G.

Choice-1: Run the risk. When a risk is run repeatedly, losses must occur over time, which affects system net output value. If the average loss (or cost) per time unit due to the hazard occurring is L_r, then the net value output from running the risk in reality will on average be $F + G - L_r$ per time unit.

Choice-2: Do not run the risk. In order not to run the risk while preserving transformation of N to U and the extra gross (before costs) gain G there are three basic choices: (a) avoid the risk; or (b) prevent the risk by investing resources; or (c) neutralize the risk with an environment coping procedure, which we assume for the present is not an open option. Avoiding the risk will involve alternative processing to transform N to U and achieve gross output F + G, and this will involve a cost (L_a) in losses per time unit. Preventing a risk involves a cost of investing additional material resources to counter the hazard; in order to maintain transformation of N to U and achieve gross output F + G, suppose investment of resources P, corresponding to L_p per time unit, either to eliminate the hazard or to eliminate the loss when the hazard occurs. Thus the net output from the transformation of N to U by a method that does not involve running the risk will be either $F + G - L_a$ in the case of risk avoidance, or $F + G - L_p$ for the case of risk prevention.

Thus for a system in an environment E_2 with a risk whose successful (in theory) running would be beneficial by producing additional gross system output G, there is a cost L_r to running the risk and a cost L_a or L_p to not running the risk. We can distinguish the following cases:

Case A. Risks which it can pay to run repeatedly : $L_r < G$ & $L_r < L_a$ & $L_r < L_p$

There is a positive extra net value output $G-L_r$ on average from running the risk. This extra output cannot be improved while maintaining the transformation of inputs to outputs by either avoiding the risk or investing resources to prevent it.

Case B. Risks it can pay to run repeatedly but can pay better to either avoid or prevent: $L_r < G$ & $((L_a < L_r)$ or $(L_p < L_r))$

There is a positive extra net output $G-L_r$ on average from running the risk. This extra output can be made even more positive, while maintaining the transformation of inputs to outputs, by either avoiding the risk or investing resources P to prevent the risk.

Case C. Risks it cannot pay to run but can pay to either avoid or prevent: $L_r > G$ & $((L_a < G)$ or $(L_p < G))$

There is a negative extra net output $G-L_r$ on average from running the risk. This extra output can be made positive or at least less negative, while maintaining the transformation of inputs to outputs, by either avoiding the risk or investing resources P to prevent the risk. [Thus, in both Cases B and C, output value V can be increased by increasing P.]

Case D. Risks it cannot pay to run or avoid or prevent: $L_r > G$ & $L_a > G$ & $L_p > G$

There is a negative extra net output $G-L_r$ on average from running the risk. This output cannot be made positive, while maintaining the transformation of inputs to outputs, by either avoiding the risk or investing in resources to prevent it. In this final case the transformation will typically never be undertaken, and need not concern us.

It seems reasonable to assume that system agents will one way or another become aware of the nature of risks in the system operating environment. Case-A risks will be run. Case-B risks will often be run and often prevented. Case C risks will be either avoided or prevented, but sometimes first run, due to unforeseen environment circumstances, and only later prevented or avoided by investing resources P.

Synthetic environments and relationship between output value, resources and risk

Most of this discussion has been for environment E_2, where total net output from running the risk in E_2 and applying resources R is $F + G - L_r$. Assume now for E_2 that $G-L_r$ is positive so that E_2 has only a risk it can pay to run repeatedly. Now recall that environment E_1 was risk free giving out F value units per unit time for resources R applied. In both environments input N is transformed to output U. It is clear that there will be an increase in output value by shifting R, from an environment E_1 with no risk and output F, to a similar one E_2 with output value $F + G - L_r$, that differs only in having a risk it can pay to run repeatedly.

But out of E_1 and E_2 we can construct an arbitrary number of synthetic environments, each with risk it can pay to run repeatedly, intermediate between the zero risk of E_1 and the risk L_r of E_2. Assuming resources R are large enough, we can do this in actual practice by taking a fraction of R and applying it to E_1 and the remainder of R to E_2. For example, suppose a such a synthetic environment E_s when the fraction is 50%. In

E_s, for the resources R/2 operating risk free, the value output will be F/2, and for the remaining resources in risky operations, it will be $(F + G - L_r)/2$ for a total of $F + G/2 - L_r/2$. Thus in E_s the risk, with resources R applied, will be $L_r/2$, and the increased output in excess of F will be $(G - L_r)/2$. We can repeat this with any fraction, so that it is clear that for such synthetic environments, net output V, or more precisely, the extra output value due to risk, will increase linearly with risk. We can even have a synthetic environment E_s where the risk of loss exceeds L_r, if we include the case where additional resources are borrowed at the cost of the risk free retrurn and applied to E_2; for example, if we borrow an additional R resources, and apply them to E_2, the output value is now $(F + 2(G - L_r))$ for the system, or $F + 2G - 2L_r$, so that the extra output for agent's resources R is $2G - 2L_r$ when the risk is $2L_r$, consistent with output value increasing linearly with risk.

It now important to grasp that, because of agent competition, in every similar but naturally occurring environment with risk $L_r/2$, the extra output for the system due to risk, with R applied, must still be $(G - L_r)/2$. From this it follows that output value must increase linearly with R and with the environment risk r_P, where it is risk it can pay to take; the output will be proportional to R when there is no risk. Accordingly we must have:

$$V = KR(1 + b_1 r_P) \qquad (2a)$$

where r_P is a measure of the class of risk which it can pay to run, which we may also call positive risk. This expression is consistent with expression (5) used in financial environments.

The above conclusion will also hold for the case where there is no competition between agents, for example the case of an agent who is a lone colonist on Mars. Assuming the agent is rational, the agent will pick out the environment E that is risk free and the environment H that has the highest extra return for the risk. For the range of synthetic environments constructed from applying a fraction of R to H and the rest to E, expression 2a must hold, since it can not pay to use any other environments.

It should also be clear from a similar analysis of a set of synthetic environments based on an environment E_4, containing only a risk it can not pay to run, that we must have

- $$V = KR(1 - b_2 r_N) \qquad (2b)$$

where r_N is a measure of the class of risk which it can not pay to run, which we may also call negative risk.

Thus in general we must have:

$$V = KR(1 + b_1 r_P - b_2 r_N) \qquad (2c)$$

where either b_1 or b_2 is zero depending on the environment. [Note that it is not possible to have non zero values for both b_1 and b_2. When both types of risk are present, a fluctuation in V due to one type of risk may well be offset by a fluctuation due to the other kind of risk, so that the result is an amalgam of the risk, the result of which will be either positive or negative risk, that is, either a positive non-zero b_1 value or a positive non-zero b_2 value. Combinations of risk have been extensively studied in finance, but the topic is beyond the scope of this paper, for example, see Markovitz's classic paper [18].]

Furthermore, in some environments, such as E_5, it will be possible to apply additional resources P (costing L_p or L_a on a per unit time basis) to prevent or avoid risk, without affecting G, with the cost increase due to L_p or L_a from P being offset by the even larger decrease in L_r (Cases B and C above). Hence, applying a similar analysis, with a set of environments based on E_5, net output value V must rise as P increases, so that the full relationship must be

$$V = KR(1 + b_1 r_P - b_2 r_N + b_3 P) \qquad (2)$$

where either b_1 or b_2 is zero, depending on the environment. Resources P are kept separate from resources R in (2) in order to have an explicit method of accounting for how risk preventing or avoiding resources P affect V independently of other resources R, and also independently of the risk factors $b_1 r_P$ or $b_2 r_N$, accounted for in (2) as if P were absent. Of course, if an investment level of P has been made in an environment, such that it can be moved with R from one environment to another, then P can be hidden in R, with (2) converted to:

$$V = HR_1(1 + h_1 r_P - h_2 r_N) = H(R+ P)(1 + h_1 r_P - h_2 r_N) \qquad (2d)$$

where now either h_1 or h_2 is zero, depending on the environment, and where the risk factors $h_1 r_P$ and $h_2 r_N$ are now accounted for assuming P to be present.

Expression (2) states that, with $b_2 = 0$ and $P = 0$, as the resources R of a system are applied to riskier and riskier positively risky environments, output value increases. However, it should be obvious that in some cases, with further environments, if the increasing risk is due to addition of negative risk, output will peak and then fall off and even go negative, as b_1 goes to zero and b_2 becomes positive. This behaviour of output with respect to risk embodied in expression (2) was noticed by Adam Smith 250 years ago [25] in the economic arena when he wrote: "The reward for assuming risk increases in proportion to the risk taken except where the risk is very great and the venture very speculative, in which case the reward for assuming risk may vanish or even be negative".

In concluding this section, readers are cautioned that if a system consists of two parallel subsytems, with outputs V_1 and V_2 such that total system output value is $V = V_1 + V_2$, where

$$V_1 = K_1 R_1(1 + d_1 r_{P1}) \quad \text{and} \quad V_2 = K_2 R_2(1 + d_2 r_{P2})$$

although the net output values are additive the risks are not, and the total system output will be governed by: $V = KR(1 + b_1 r_N)$
where the risk for the total system is a statistically determined combination of the separate risks for the two subsystems; this combined risk can even be zero, in cases where a downward output fluctuation in one susbsystem is always counterbalanced by an upward fluctuation in the other subsystem; this topic is beyond the scope of this paper, but is of great importance in financial systems where the aim is to reduce risk without reducing the expected output V; see [18].

A final point to note is that in financial systems with moderate risk b_2 is almost always zero, because fast acting market forces ensure an extra return for taking (moderate) risk; in contrast, in non financial systems, although market forces will eventually make b_2 zero, initially, with new systems, due to poor design or lack of understanding of the true nature of the environemnt, b_2 may be positive.

The relationship between MD and SD-risk and output and resources

It remains to demonstrate that the risk parameters in (2) can be treated as either SD-risk or MD-risk. An example should suffice. Suppose a system with resources R =

1000 value units. Suppose it operates in a risk free environment E_1 transforming inputs N to outputs U with a regular net output value F equal to 200 value units per unit time. Suppose now we shift the environment to a similar one in which the system still transforms inputs N to U but for a randomly fluctuating net output value. Assume the output is 250 on average, but will have randomly fluctuating values 350, 150, 150, 350, ... of 350 or 150. This means an expected or average value output of 250 and an MD-risk r of 100. The output per unit time if the risk is run but the hazard (a downward fluctuation) never occurs will therefore be 250 + 100/2 or 300 on average per time unit. Thus F + G must equal this 300, so that G = 100; the risky environment contributes 100 per time unit to output F = 200, the output of the risk-free environment, if the hazard does not occur. But the hazard does occur repeatedly, and the output can in reality fluctuate downward to 150, one time period in two, for an average loss of 100/2 or 50, which must be L_r. (Since L_r is an average loss due to risk, it equals half the MD-risk of 100, as demonstrated in the analysis deviation risk measures earlier.) Hence $L_r = r/2$, and is thus also a measure of risk. Hence the net output must be

$$F + G - L_r = F + G - r/2 = 200 + 100 - 50 = 250$$

which is also the expected value. We can therefore treat $F + G - L_r$ as the expected output, and L_r as half the MD-risk, when the output due to risk is fluctuating. In this case of MD risk of 100, since G is 100 and $L_r = 50$, the value $G-L_r$ is positive, and shifting to E_2 involves a risk it can pay to run repeatedly, as it increases the expected value of 200 to 250. An analysis in terms of SD-risk is very similar but less tidy in this case.

For the environments E_1 and E_2, other similar environments and synthetic environments constructed by applying the resources R = 1000 partly to E_1 and partly to E_2, expression (2a) will apply to systems converting input N to output U. Substituting appropriate values for the constants, the expression becomes:

$$V = 0.2R(1 + 2.5\, r_P) \qquad (2e)$$

if the risk variable r_P is a mean deviation measure (MD-risk) expressed in per unit values. (MD-risk of 100 is equivalent to an r_P value of 0.1 (per unit) of 1000). In (2e) b_2 is zero and $b_1 = 2.5$.

Suppose now, to show how the additional rsource quantity P fits in, that if resources P of 100 value units are added, but portable from one environment to another, with the result that in the environment above the downwards fluctuations of V are reduced by 50 units, that is, V now fluctuates between 350 and 200, instead of 350 and 150. And suppose further, for the sake of simplicity of presentation, that P only depreciates to zero over many thousands of time units, so that the cost of P per unit time can be neglected. It is clear that the expected value of V is now 275, and the total resources R are now 1000 + 100 = 1100. A new analysis identical to that above gives us an MD risk of 75 and:

$$V = 0.182R(1+ 5.5r_P) \qquad (2f)$$

corresponding to expression (2d), and where in this environment r_P is 0.0682 per unit of 1100. However, in (2d), as already pointed out, the effect of P is hidden. If we modify (2e) to

$$V = 0.2R(1 + 2.5\, r_P + 0.25P) \qquad (2g)$$

corresponding to expression (2), the effect on expected output value V of adding resources P becomes explicit. However, it must be noted that in (2g) the factor $2.5r_P$ respresents the contribution of risk to V as it would be if P were not present. Expressions

(2f) and (2g) are equivalent, and, with P = 100, both give a value for V of 275. However in (2g), the original environment risk of $r_P = 0.1$ is used, whereas in (2f) the new risk $r_P = 0.0682$ of the environment with P = 100 present is used, to arrive at V = 275.

3.0 Resources versus Monitoring & Sharing Complexity

This part of the Fundamental Systems Hypothesis states that for a system with a given environment coping procedure, and given resources R, output V can be increased by increasing the monitoring complexity c_c in the coping procedure, and also increasing the complexity resource-sharing procedure time t_s (for resources deployed by the system) in accordance with

$$V = KR(1 + M(c_m) + M(t_s)) \qquad (3)$$

To show that (3) holds generally we need to solve three problems. First we need a metric for the monitoring complexity of a coping procedure. Second, for a given level of system output value we need to show that increasing monitoring procedure complexity will increase output value. Finally we need to show how increasing complex resource-sharing activity will increase output value.

In approaching the first problem, we observe that a system can benefit from a complex procedure that deals with the system environment only if the complexity of the procedure is oriented directly towards coping with that environment. The degree to which a procedure copes with its environment corresponds to the degree to which it can extract positive net value from it. But for a system operating in real time the environment may unfold with a degree of unpredictability or randomness. Consequently the complexity in the environment-coping procedure must directly depend on the procedure's ability to monitor and detect regularity in the unfolding environment, as this will enable it to predict the unfolding environment, and so enable it to exploit opportunities and avoid pitfalls.

An effective complexity measure in relationship to algorithmic information content

Consider any system operating in real time and confronted with an unfolding environment that can be characterised by a continuous data stream of bits, ones and zeros. For example, if it is a mobile robot system, the data stream consists of sensor data, if it is the paging supervisor of a virtual operating system, the data stream consists of a history of page access data, or if it is a shortest-job first scheduler in an operating system, the data stream consists of the history of cpu burst data, or if a deadlock avoidance monitor using the banker's algorithm in an operating system, the data stream consists of allocation and request data, or if we have a financial management system, the data stream comes from the markets and financial reports, and so on. Data that has already arrived constitutes a history of past environment unfolding, i.e. the historical data string.

We can characterize the data stream in terms of its Algorithmic Information Content (AIC). The AIC is simply the length of the briefest algorithm, which, given the data string already stored in memory, can print the string. AIC is not a complexity measure.

If the historical data string is perfectly random, then the AIC will be the length of the data string near enough, since there are no string compression possibilities; the algorithm to print the string must be:

Print ('01100110101000101 ...01101').

If the data string is perfectly regular, such as '01' repeated n times, the algorithm will now be brief and the AIC will be small:

Print ('01') *n times*.

If the data string is random but with some order, or regularities, the algorithm can incorporate string compression techniques, in which case the AIC will be intermediate between that for the regular string and that for the random string.

However AIC does not tell us how complex the historical data string is. An useful idea proposed by Gell-Mann and others [10, 11] is that the most appropriate complexity measure for the string is an effective complexity measure c that must equal, or at least be proportional to, the length of a (concise) specification of the regularities or order in the string. Accordingly, c must be very small or zero both when the AIC is zero (completely regular string) and when it is at a maximum (completely random string with no compression possible). But most importantly, c must have a maximum (or perhaps several maxima) at some point(s) intermediate between minimum AIC and maximum AIC.

It would therefore appear that the specification-length measure of effective complexity for an historical data string must in some way be related to the complexity of a coping procedure to monitor and exploit an unfolding environment signaled by a historical data string of effective complexity c. We shall argue later that if the effective complexity of the incoming data stream is c, then the monitoring complexity in any relevant coping procedure needed to monitor the regularities in the string must also be c or proportional to it. If c is zero for the string, then the future is either completely predictable or completely unpredictable (random) so that the monitoring complexity of any relevant coping procedure must be zero, with no possible contribution to increased system output value.

Incoming time-function data stream and the Hurst constant H

Rather than consider effective complexity of an historical data string in relation to AIC, the author has found it somewhat more enlightening to consider c in relation to the Hurst constant of an equivalent time function. [Hurst was a British engineer who studied time series of water levels on the Nile around 1900, in an effort to design (complex) systems to deal with flooding]

Instead of considering the historical string as a bit sequence, consider it as multi-pulse time function B(t), randomly generating a sequence of values 0.5 & -0.5 corresponding to bits 1 & 0. Conventionally, if the values of B(t) are the values of a random variable G_i (a fair coin toss), the sum (or integral) of B(t) to time t will be a random walk time function S(t) about origin zero. If the values of B(t) are still 0.5 and -0.5 but only partly random, the integral of B(t) to time t will yield a partial random-walk time function S(t). Thus we can study effective complexity of the historical summary function S(t), instead of the underlying historical bit string B(t). This is more useful, since the regularity in a partial random walk can to some extent be characterized in terms of the

Hurst constant H for the function, or more precisely, by the degree to which H differs from 0.5, as will be explained presently.

Now suppose $s_i(n)$ is the sum of n consecutive pulse amplitude values B(i+1), ...B(i+n) beginning at time i+1; where n is small compared to the total number of single-pulse functions summing to S(t) Suppose now a large collection of such sums $s_i(n)$ for constant n but beginning at different i values. Applying the Binomial theorem to this collection of $s_i(n)$ values:

$$\text{StDev}[s_i(n)] = (pqn)^{0.5}$$

where p and q are the probabilities for the G_i. Since in S(t), n is proportional to the time T between pulse B(i) and B(i+n), and since

$$s_i(n) = S(i+n) - S(i)$$

it follows that

$$\text{StDev}[S(t + T) - S(t)] = KT^{0.5}$$

or the well-known result that in a random walk the standard deviation of the increments to S(t) over time intervals T is proportional to the square root of T (to go twice as far on average in magnitude S, it takes 4 times as long). The mean of the increments is zero. The above expression is fundamental and can be rewritten:

$$\text{StDev}[S(t + T) - S(t)] = KT^H$$

where H is the Hurst constant, equal to 0.5 for a random walk or Brownian motion in one dimension.

. However, for many time series S(t) representing an unfolding environment in the real world, the StDev[S(t+T) - S(t)] is not proportional to $T^{0.5}$, but to T^H where H is <0.5 or >0.5, with increment mean equal to zero. Mathematically, a time function S(t) with H not equal to 0.5 is a generalization of a random walk, often referred to as a fractal time function or fractal Brownian motion time function [6, 12, 15, 16, 17, 28] .

A value of H close to 0.8 seems to hold for many series, including economic time series, river heights over time, terrain height as measured over time by airborne telemetry systems, and so on [16, 28]. A value close somewhat about H = 0.0 occurs in music [29] and many physical systems [9].

Functions with $-0.25 < H < 0.25$, including music, are usually called 1/f noise, since the spectral density [8, 20, 28] varies with 1/f. Spectral density at frequency f estimates mean square amplitude fluctuations at frequency f. If spectral density of a time function is K/f^B, it can be shown that B = 1 + 2H, so that spectral density for music (H = 0) must vary with 1/f, which means that the amplitude of high frequency fluctuations is lower than that of lower frequency fluctuations. Unfortunately, despite much research, the mechanisms behind 1/f noise in nature are not understood [29]. For a random walk spectral density varies with $1/f^2$.

Fractal time functions with H <> 0.5 are poorly understood. It is generally not possible to give a simple set of rules that will cause one to be generated, like the simple rules for the random walk case with H = 0.5. We can only simulate these time functions from a sum of generalized single-pulse time functions [28]. Simulations throw little light on any physical mechanism behind fractal Brownian functions, however.

Time functions with constant H over time and zero mean increment can be superimposed on time functions whose mean increment is non zero, that is, regular growth or decay time functions. An example would be a function with H=0.8 and mean increments zero superimposed on f(t) = kt. The new function would still have H = 0.8, but non-zero mean increments. Since the additional regularity due to the superimposition typically adds zero to a complexity specification, we ignore this class of function and restrict our attention to functions with mean increment zero, that is where S(t) always fluctuates about zero.

We can classify all such time functions in terms of H.

1. Self-affine or fractal time functions, H is constant.

Examples are random walks with H= 0.5, or functions with H <> 0.5. Since the standard deviation of function increments over time T is proportional to T^H, these functions are statistically similar, that is, statistically fractal-like or self-affine, at all levels of scaling for S(t) and T. Such functions are idealistic and probably do not exist in nature. As we shall see, for H > 0.5 the functions are trend prone, for H < 0.5 they are reversal prone.

2. Pseudo self-affine functions where H is constant for T-values up to a maximum of T.

These functions are self-affine over a finite range of scaling. If T is very large we can treat H as a constant. Functions of this kind are common in nature.

3. Time functions where H varies slowly with T.

Functions of this kind do not appear to be very common. However, for short periods T, they often have constant positive H, but as T gets very large H typically falls. Time functions derived from underlying bit streams generated by pseudorandom number generators are ultimately of this kind.

4. Time functions where H oscillates over T.

With these functions, for one period of time D of random length, H appears to have one constant value for T < D, but in a succeeding period of random length, H has another value, and then, in the next period, H reverts to close to the original value, and so on. Such functions, although uncommon, since they are both trend prone and reversal prone, are extremely difficult to specify regularity for, and recently there have been attempts to use neural nets that can learn in real time [21] in this area .

For our purposes the really important aspect of these functions is that when H > 0.5, S(t) will be random but trend prone, such that if S(t) increases over a time period d, there is a greater probability of an further increase over the next d time units than a decrease, whereas with a random walk and H = 0.5 the probabilities of increases and decreases are always equal. Also with H < 0.5, S(t) is random but is reversal prone, such that if S(t) increases over a time period d, there is a greater probability of a decrease (a reversal) over the next d time units than an increase. We will demonstrate the truth of this presently, but if the above is true then one would expect a positive autocorrelation coefficient for S(t) when H > 0.5, a negative one for H < 0.5 and a zero one for H = 0.5.

H as indicator of trend or reversal propensity of a time function

To demonstrate that (H-0.5) is a measure of the trend-prone or reversal prone nature of the time function S(t), suppose W_i is any increment of S(t) over time T, where W_i in turn breaks down into T successive increments I_{i1}, I_{i2},... I_{iT} of S(t), with one increment per unit time. If we use notation < > to denote the average of the of sum of many similar expressions, each for a different increment W_i, the variance of the W_i over time period T will be $<(W_i)^2>$ or:

$$<I_{i1}^2 + I_{i2}^2 + \ldots I_{iT}^2 >$$
$$+ \ < I_{i1}I_{i2} + I_{i1}I_{i3} \ldots + I_{i2}I_{i1} + I_{i2}I_{i3} \ldots>$$

For a random walk the elements $I_{ik}I_{im}$ will be distributed randomly about 0, the correlation coefficient for I_{ik} versus I_{im} will be zero, so that the variance thus reduces to

$$<I_{i1}^2> + \ldots <I_{iT}^2>$$
$$= T<I_{i1}^2> \quad = TK^2$$

so that the standard deviation is $KT^{0.5}$. Thus H=0.5, when successive increments of S(t) are uncorrelated, that is, if the probability p of the next increment being positive, or negative, is 0.5, as derived earlier from the Binomial Theorem.

Now suppose that for increment I_{ik} in unit time period that the next increment $I_{i(k+1)}$ has probability p of having the same sign as its predecessor. In that case, the term

$$< (I_{i1}I_{i2} + I_{i1}I_{i3} \ldots + I_{i2}I_{i1} + I_{i2}I_{i3} \ldots)>$$

above must have a fraction p - (1-p) or 2p-1 of its elements of the form $I_{ik}I_{i(k+1)}$ or $I_{i(k+1)}I_{ik}$ that do not cancel and which are all of the same positive sign for $p > 0.5$ and all of the same negative sign for $p < 0.5$. The variance therefore reduces to

$$<(I_{i1}^2 + I_{i2}^2 + \ldots I_{iT}^2) >$$
$$+ 2[<(I_{i1}I_{i2} + I_{i2}I_{i3} + I_{i3}I_{i4} + \ldots + I_{i(T-1)}I_{iT})>](2p-1)$$

$$= TK^2 + 2K^2(2p-1)T$$

Hence the standard deviation must be:

$$K[(4p-1)T]^{0.5}$$

If we let p = 0.5 + h, the standard deviation becomes, to a fair approximation

$$K[(1 + 4\,h)T]^{0.5}$$

When p is 0.5, or h = 0, for the case where a successor increment in any unit time period is as likely to be positive or negative, this standard deviation reduces to $KT^{0.5}$. But where $p > 0.5$, the standard deviation exceeds $KT^{0.5}$, and is less than $KT^{0.5}$ when p is less than 0.5.

The above result comes from assuming there is a probability p that an increment of W_i has the same sign as its successor. But this will mean that there is also a (much smaller) probability (p_1) that an increment of W_i has the same sign as the increment immediately after the successor. This will have a small affect on the evaluation of

$$< (I_{i1}I_{i2} + I_{i1}I_{i3} \ldots + I_{i2}I_{i1} + I_{i2}I_{i3} \ldots)>$$

and we can get a better result for the standard deviation by taking it into account.

It is easy to show that

$$p_1 = p^2 + (p-1)(p-1) = 2p^2 -2p + 1$$

Since therefore $\quad (1-p_1) = 2p - 2p^2$

we must have a fraction $\quad (4p^2 - 4p + 1)$

of terms of the type $I_{i1}I_{i3}$ or $I_{i3}I_{i1}$ that do not cancel, that is, an additional contribution to the variance of:

$$2[<(I_{i1}I_{i3} + I_{i2}I_{i4} + I_{i3}I_{i5} + \ldots + I_{i(T-2)}I_{iT})>](4p^2-4p+1)$$

that is: $K^2(4p^2-4p+1)T$, or $K^2(4h^2)T$

allowing for the fact that there are half as many terms as previously. This contribution is clearly minor but not negligible. Contributions from other terms involving more distant successors will be negligible, however. Adding this contribution to the standard deviation, it becomes:

$$K[4p^2T]^{0.5} \text{ or } 2pKT^{0.5} \text{ or } (1+2h)KT^{0.5}$$

We can rewrite the standard deviation as KT^H where $H = 0.5 + X$ where

$$T^X = 2p,$$

so that $X = \ln(2p)/(\ln(T))$,

and where, as is typical of this kind of analysis, this H is not constant with T, and we still get a random walk at large T; X is also zero at $p = 0.5$. Still, it is simpler to note that in $2pKT^{0.5}$, as T increases, the factor 2p becomes less important than the factor $T^{0.5}$.

Thus when $p > 0.5$ the time function S(t) must be trend prone and the Hurst constant H must exceed 0.5 provided T is not too great, since the inherent trend due to p will drive S(t) further from 0 in either direction over a time period T than would otherwise be the case, thus giving a higher standard deviation (proportional to p) than otherwise for increments of S(t) over T. Similarly, when $p < 0.5$ the function S(t) is reversal prone, which will make the standard deviation (still proportional to p) less than otherwise, so that $H < 0.5$.

This simple analysis demonstrates clearly the fundamental nature of the relationship between H and autocorrelation coefficient for the time function - when $H > 0.5$ the current trend is likely to continue, but is likely to reverse if $H < 0.5$. Unfortunately, there seems to be no simple relationship between H and autocorrelation function. However the fact of the nature of the relationship is enough for our purposes.

It should thus be clear that there are two types of regularity occurring in quasi-random time functions, namely order of a trending nature in functions where $H > 0.5$ and order of a reversal nature in function with $H < 0.5$. As H increases from 0.5, trending order increases from zero, and the function becomes more and more predictable.

However, as H decreases from 0.5, reversal order increases from zero. When H is 0.0, the standard deviation of the function increments is constant with T, so that the function is highly reversal ordered. This is also 1/f noise, discussed above. When H is -0.5, we have a white noise function [8, 22], of which the random walk is the integral. The spectral density for white noise [8, 22, 28] varies with $1/f^0$, so that all frequencies are equally present. A white noise function can be taken as being the values of a random number generator, and is highly disordered. It is the derivative of a random walk.

The zero-regularity in a time function with H=0.5 seems to correspond to Gell-Mann's zero effective complexity for a random string with maximal AIC. The fractal function with H=0, that is, standard deviation constant with time period, is very reversal prone, and contains a mixture of randomness and predictable reversibility, like music. It probably corresponds to low effective complexity for a highly ordered string with minimal AIC. But maximum effective complexity for a random string with inherent regularity and medium AIC has two possible correspondences: to a maximally complex reversal-ordered time function with $0.0 < H < 0.5$ and to a maximally complex trend-ordered time function with $H > 0.5$. Currently it does not appear possible to state at what, if any, specific H values these complexity maxima will occur. All the current state of knowledge allows us to state with confidence is that when the trend (or reversal) order of a fractal time function is maximally complex it will take a longer specification to specify its regularity than that of any other functions with lesser or greater H, that is, with greater or lesser AIC equivalent. In addition, and most importantly, with this regularity goes predictability.

Effective monitoring complexity of a coping procedure monitoring a time function

A coping procedure that exploits the trend or reversal propensity of a time function whose effective complexity is c must contain the equivalent of a sequence of pairs of conditional (or prescriptive) imperatives of the form:

If <order specification-type-S_{A1}> then <exec task-type-A>;
If <order specification-type-S_{A2}> then <stop task-type-A>;

If <order specification-type-S_{B1}> then <exec task-type-B>;
If <order specification-type-S_{B2}> then <stop task-type-B>;
...

In practice the coping procedure must have a monitoring procedure for detecting the order specifications for detecting regularities in the unfolding environment.

With completion of a task-type-A there is either an output gain G_A, where (S_{A1},S_{A2}), in detecting regularity type A, signals an opportunity, or there is an output loss avoidance G_A, where (S_{A1},S_{A2}) signals a pitfall; however with execution of task-type-A, G_A is not certain, there is only a probability p_A of its occurrence, with a probability $(1\text{-}p_A)$ of a loss L_A instead. Similarly for task types B, C etc. following detection of regularity types B, C, etc.

The monitoring procedure may be said to be saturated with monitoring complexity if all regularity types A, B, C ... inherent in the time function being monitored are coded for in the procedure's order-specifications. For a saturated monitoring procedure, it follows that the sum of the lengths of S_{A1}, S_{A2}, S_{B1}, S_{B2}, ... must be equal to, or proportional to, the effective complexity c of the time function, since these specifications reflect all the regularity types A, B, C ... in the time function. If the procedure is unsaturated, that is, not all regularities in the time function are coded for, we define the sum of the lengths of S_{A1}, S_{A2}, S_{B1}, S_{B2}, ... as the effective monitoring complexity c_m of the monitoring procedure. Adding additional conditional-imperative pairs to the monitoring procedure, with order specifications corresponding to those in the function not already coded for, will then increase the procedure's effective monitoring complexity up to its maximum of c.

To derive the relationship between output value and monitoring complexity, we note that each regularity type J detected in the unfolding time function by (S_{J1}, S_{J2}) will have a specific frequency of occurrence; some regularity types will occur very often, others occur only rarely. Futhermore the occurrence of one regularity type J may give rise to a large gain (or prevent a large loss) G_J when correctly exploited, whereas another may give rise to only a small gain or prevented loss. Finally, one regularity type J may give rise to a high probability p_J of the associated gain or prevented loss G_J taking place, whereas another may have only a small probability.

It is useful at this point to define a *monitoring complexity efficiency coefficient* e_m. There will be an e_m value associated with each of the regularity types A, B, C occuring in the time function; for any regularity type J, e_m is the the contribution to V over a specific time period of the algorithm to successfully detect J, divided by the length and thus complexity of that algorithm. This means that for the unfolding function over a given period of time, we can order the complexity types A, B C, ... in order of their decreasing monitoring-complexity efficiency coefficients. Thus those that cummulatively over the period give rise to the largest contribution to V for the least complexity appear first (say type A), by virtue their relative simplicity and combination of high frequency, high gain or prevented loss G_A and high probability p_A of G_A; those with next largest ratio of contribution to V (type B) for length of detection algorithm appear next, and those with the lowest ratio appearing last. We refer to this function of cummulative V versus cummulative c_m, that is, $V = KRM(c_m)$, where monitoring complexity is added in order of decreasing e_m, as the *monitoring complexity schedule*.

We can assume that a rational designer would pregressively add monitoring algorithms to the system in the order of the monitoring complexity schedule or decreasing e_m, that is, the simplest algorithm with the most effect on V would be added first, and so on, until the point is reached in the limit, where the monitoring procedure is saturated, addition of a very complex and lengthy algorithm would have negligible impact on V. Thus the relationship between monitoring complexity and V is given by an expression

$$V = KR(1 + M(c_m))$$

where M(x) is a function of the general form:

$$M(x) = m(1 - e^{(-kx)})$$

so that when the monitoring algorithm is saturated $V = KR(1+m)$ so that $V_{max} = KRm$, and where far from saturation, where only the simplest detection algorithm has been added to the system, $V = KR(1 + mkc_m) = KR(1 + a_1c_m)$ so that $a_1 = mk$. The monitoring-complexity efficiency coefficient e_m is always dV/dc_m and is therefore zero when c_m approaches infinity.

For an unsaturated monitoring procedure J with effective monitoring complexity c_J, and code for regularities A, B, C ...J, in order of decreasing e_m, the value contribution to v to V will be:

$$v = KRM(c_J) = [p_AG_A - (1-p_A)L_A] + [p_BG_B - (1-p_B)L_B] + ... + [p_JG_J - (1-p_J)L_J]$$

Effective resource scheduling & utilization complexity

Every system involves one or more processors, each engaging in a sequence of processes, tasks or activities, not necessarily contiguous in time, that each utilize resources. For a given processor for a specific process, a set of resources will be needed, and a different set for the subsequent or prior tasks. As a result, in every system, even the simplest involving only one processor, each of the individual resources required by a processor, and the processor itself considered as a resource, is unlikely to be in use 100%

of the time. As a result, overall system resouce utilization will typically be much less than 100%.

However, it is easily observed that with any system involving a single processor and multiple activities, the required resources can be reduced, or more fully utilized, to give greater output value, if the processor also engages in a usually complex and time consuming management activity designed to share resources between activities. If we consider a human as such a processor, many examples, often amusing, are apparent. For example, a person with a two-room apartment - living room and bedroom, on noticing the "waste" of resources in the fact that when sleeping in the bedroom the living room is not in use and vice versa, decides to live in a one-room apartment instead; this saving in space resources has to be paid for by a daily time consuming and complex process of converting the bedroom to a living room in the morning and the converse in the evening. Another example, involving saving of processor resources, is the dentist who, on noticing that much valuable time is wasted in waiting on physical events, multiplexes the time spent between many patients, much like a multiprogramming uniprocessor operating system; the price paid for the better use of the dental processor resource is the time consuming management of the multiplexing activity. A simple computer example is the case of a single processor that is used repeatedly to search a large file, with help of an index stored on the disk. Since the file and its index are never in use together, disk space resources can be saved by eliminating the index and using a sequential search of the the file instead, at a cost of much increased file searching activity.

Examples with multiple processors are also ubiquitous. The classic example involving two human processors is the case of husband and wife with multiple diverse daily engagements requiring travel; they have only one car to save resources but pay for this saving with a complex management procedure to schedule and share the single car between them. Acquiring a second car (doubling of resources) instantly eliminates the need for this complex procedure. In the computer system arena, a major function of multiprogramming multiprocessor operating systems is to significantly increase computer resource utilization, in terms of giving the best value for the resources involved, but this is achieved at the cost of a time-consuming resource management and sharing procedure, the operating system.

Often the required time-consuming procedure required when resources are saved is not considered very carefully and even neglected, especially in the computer system engineering arena. The classic example is the saving of two bytes of memory resource per date in COBOL programs, without making the required payment in terms of the procedure (YEAR2000-PARAGRAPH and IF-statement omitted) for all program executions before and after 2000. Another example is the reuse of objects in object oriented programming, which has to be paid for in the routine execution of more complex programming procedures. Another one is the use by airlines, considered as systems, of very complex seat management procedures, partly computerized, for maximizing the utilization of their major resource: seat miles, and in that way increasing output value.

There is a basic Law of Nature at work here. Essentially it says that an agent can save necessary resources in a system only at the price of routine system execution of additional complex resource-sharing procedures. Many people are aware of the operation of this Law in common-sense activities where a human is considered as a processor, and are aware of the old adage: "time or money" (time-consuming procedures or resources), or the familiar resource-time or space-time trade-off. Awareness of the unforgiving nature of this Law in general in complex systems seems to be uncommon, however.

One thing is clear: the source of the saving in resources by means of complex current resource-sharing procedures is the fact that, in any system, resource utilization is less than 100% and that by sharing resources among processes and processors, by means of complex current resource-sharing operations, this utilization can be increased. However, it cannot be increased indefinitely - only to the point of 100%. And we assert that as the utilization aproaches 100%, each increased 1% gain requires greater complex resource-sharing procedure activity time than the previous 1%. This means that we can define a *resource-sharing efficiency coefficient* e_s, which is the ratio of increase in value output V to increase in complex current resource-sharing procedure activity time (t_s) at constant resources R. For a given resource level, the value v contributed to output V as a function of the level of complex current resource-sharing procedure time (t_s) activity, at constant resources R, is the *resource-sharing complexity schedule* for the system, that is:

$$v = KRS(t_s)$$

where complex current sharing procedures are assumed to be added in order of decreasing e_s.

We can assume that a rational designer would progressively add complex resource sharing algorithms to the system in the order of the resource-sharing complexity schedule or decreasing e_s, that is, the simplest sharing procedure with the most effect on V (highest e_s) would be added first, and so on, until the point is reached in the limit, where resource utilization is approaching 100% and addition of even a very complex and exceedingly time-consuming resource-sharing procedure would have negligible impact on V. Thus the relationship between complex resource-sharing procedure time t_s and V must be given by an expression

$$V = KR(1 + S(t_s))$$

where S(x) is a function of the general form:

$$S(x) = s(1 - e^{(-kx)})$$

As a result, when resource utilization is 100%, V = KR(1+s) so that addition v to value V when resource utilization is 100%, is given by $v = V_{100\%} = KRs$, and where far from 100% resource utilization, where only the simplest resource-sharing procedure has been added to the system, $V = KR(1 + skt_s) = KR(1 + a_2t_s)$ so that $a_2 = sk$. The resource-sharing efficiency coefficient e_s is always dV/dt_s and is therefore zero when t_s approaches infinity.

Hence the expected output must generally depend linearly on the level of resources R, the monitoring complexity schedule $M(c_m)$ and the resource-sharing complexity schedule $S(t_s)$ in accordance with

$$V = KR(1 + M(c_m) + S(t_s)) \qquad (3)$$

which far from monitoring complexity saturation and 100% resource utilization, can be approximated by a linear function:

$$V = KR(1 + a_1c_m + a_2t_s) \qquad (3a)$$

4. The Relationship between Coping-procedure Complexity, Risk and Resources.

Since we have now shown that

$$V = KR\,(1 + b_1r_P - b_2r_N + b_3P) \qquad (2)$$
$$V = KR(1 + M(c_m) + S(t_s)) \qquad (3)$$

for any relationship between risk and complexity, we can have either

$$V = K\,R\,(1 + b_1r_P - b_2r_N + b_3P)(1 + M(c_m) + S(t_s)) \qquad (6)$$

or

$$V = K\,R\,(1 + b_1r_P - b_2r_N + b_3P + M(c_m) + S(t_s)) \qquad (1)$$

We can use the risk analysis in Section 2 to show that (6) must be wrong. Suppose an environment E_4 in which there is a large risk r_N it can not pay to run (but can pay to avoid), so large that with application of

$$V = K\,R\,(1 - b_2r_N + b_3P) \qquad (2)$$

net output value V is forced negative. In the analysis in Section 2 it was noted that a risk could be eliminated or much reduced by a relevant coping procedure, as well as by prevention resources P, but this was not further included in the discussion. With expression (6), no matter how we increase monitoring complexity c_m or resource-sharing procedure time t_s, V will stay negative. But risk can be mitigated by using a complex coping procedure (or can be compensated for by better resource utilization), so that $M(c_m)$ (as well as $S(t_s)$) ought to behave in the expression like P, as far as mitigating (or compensating for) risk is concerned. It follows that (6) must be wrong, so that expression (1) must be correct.

Hence, for a given level of resources R, expected value V relates to risk and complexity by:

$$V = K\,R(1 + b_1r_P - b_2r_N + b_3P + M(c_m) + S(t_s)) \qquad (1)$$

or more approximately, far from monitoring complexity saturation and far from 100% resources utilization:

$$V = K\,R(1 + b_1r_P - b_2r_N + b_3P + a_1c_m + a_2t_s) \qquad (1a)$$

where, it should be remembered, either b_1 or b_2 (or both) must be zero.

6. Concluding Remarks

It is important to understant exactly just what the Fundamental System Hypothesis in expression (1) is tells us. The following should help.

Suppose a system with resources R could be operated in a variety of similar environments E_1, E_2, ... but with differing risk characteristics, as measured by the level of positive risk (risk it can pay to run repeatedly) r_P and negative risk r_N (risk it can not pay to run repeatedly but can pay, or at least not pay by less, to prevent).

First of all, for the system in any environment E_n, expected output value V increases with increased resources R applied. This is merely stating that if you double a system, you essentially double the value output , so that R can be taken as a measure of system capacity. (However, see Appendix 1 for the case of isolated environments.)

Next it says that for a given R, if the agent directing the system shifts it from one environment to another with higher positive risk r_P, output V will also be increased, and not only that, wilkl increase linearly with r_P. If the system is shifted to an environment with negative risk r_N, expected output V will fall and may even be driven negative.

It then goes on to say that for a system in a given environment with certain inherent risks, typically positive risk, but occasionally negative risk, losses due to the hazards associated with such risk actually taking place can be reduced by the agent without reducing the gross gain due to then. This is accomplished if the agent increases the resources P devoted to preventing these risk, so that in consequence output value V will increase with increased applied risk-preventing resources P. P will normally be obtained from outside the system; such resources P may be considered as a special category of resources separate from R. Nevertheless, total resources utilized by the system will of course be P + R, and over time, a specific P will normally be integrated into R

It also states that for a system with a given environment with fixed risks, typically positive risks, output can be increased by increasing the monitoring complexity of the coping procedure (increased r_m) provided there is predictability in the environment (that is, the unfolding time function is not a random walk.) A simple financial example would be funds placed in long-term government bonds. The output value would be improved by a monitoring procedure that could signal the market swings in advance. However, in practice bond fluctuations appear to be very close to a random walk so that the monitoring complexity would be rather great and the additional output rather small, so that achieving significant improvement in V in this case, as is well known, would be difficult, if not impossible.

Finally, it says that for system with a given environment E_n with fixed risks, either the usual positive risks r_P, or negative risks r_N, or with zero risk, for a given R, and for a coping procedure where monitoring complexity c_m is zero or non zero, output value can be further increased if the agent increases the complex resource-sharing procedure time t_s, for example according to :

$$V = K\,R(1 + S(c_s)) \qquad (7)$$

This phenomenon is easily observed in computer systems, where it is common practice to increase system throughput by increasing complex resource-sharing procedure time. Obvious examples following expression (7) are multiprogramming operating systems [26] and file and database systems [1].

For specific classes of environments, the constants and variable parameters in expression (1) can be reduced to numbers and measurable quantities. Thus the Fundamental Systems Hypothesis can be subject to experimental verification. In the field of financial systems much work has already been done that confirms its validity with financial subsystems [3, 24]. In other areas, experience with complex systems points to the veracity of Hypothesis. The reader who has taken care to fully grasp the Hypothesis will no doubt be able to see that it fits with his or her own experience of practical functioning systems, especially real-time systems.

Note on the application of the Hypothesis to biological systems

Finally, we remind the reader that the systems to which the Fundamental Systems Hypothesis applies are human-agent-directed systems, that is, systems created and directed by humans for the purpose of generating value to humans. We have not considered naturally occurring self-reproducing systems, that is, biological organisms [10]. The major difference is that with a biological system in the wild the net value V output largely remains with the system as an increase or decrease in resources R (and perhaps P), except when reproducing; the net value output V will be positive while the organism is growing, zero in maturity and negative in decay; it will also be positive when the organism is reproducing itself, the V being transferred to the resources of child biological systems. But as in agent directed systems, net value output by a biological organism depends on its resources, risks in its environment, its environment monitoring mechanisms (monitoring procedure equivalent) and its resource-sharing mechanisms (like a hand that is used both for manipulation and support in primates) for responding to its environment. Thus, if the Hypothesis is applicable to biological organisms, and the author can find no reason why it should not be, at least to a good approximation, then the Hypothesis should also throw light on biological evolution possibilities. This subject, of course, lies well outside the scope of this paper.

Appendix 1

Because on occasion systems operate in an isolated "Robinson Crusoe" environment, where price per unit output may be sensitive to supply of units, an expression for K in terms of price per unit and supply/demand parameters is required for completeness. Consider a collection of systems, all producing items of the same type, physical or informational, e.g. apples or copies of a report. Let supply schedule be $p = S + N/s$ and demand schedule be $p = D + N/d$, where p is price per unit, N is the sum of units produced by all systems, s and d, are respectively the elasticities of supply and demand, and S and D are constants, namely the limit prices where potential supply of units and demand for units is zero. Solving, the price at which supply matches demand is $P = (Ss + Dd)/(s+d)$ at which point total units produced is $N = sd(D-S)/(s + d)$. But since for a specific additional system, units produced n in must be proportional to resources deployed R, so that total units is $N + n = N + kR$, then the value of the production by that specific system must be $V = n(sS + dD)/(s+d) = k(sS + dD)/(s+d)R$. Hence the value for the "constant" K used in expression (1) is

$$K = k(sS + dD)/(s + d)$$

Hence, technically, K is effectively constant with changing resources R for a system only if the elasticity of supply s for the collection of systems is effectively unchanged., or only slightly changed, with changing R for a specific system. Clearly if we significantly change resources R in a specific system and so change N for all systems operating, at a given price per unit, then, then since N always equals $sd(D-S)/(s+d)$ no matter how much resources are added, it is supply elasticity s that must have changed. But if the change in N due to the specific system is small, then then s will effectively unchanged and so will K; as pointed out above, this will be the case for systems operating in a large economy - but not in a tiny isolated one.

References

1. Bradley, J., "File and Database Techniques", Holt, Rinehart and Winston, New York, 1981.

2. Beaumont, G. P., "Probability and Random Variables", Ellis Horwood Limited, 1986.

3. Brealy, R. and Myers S., "Principles of Corporate Finance", McGraw-Hill, N.Y., 1981.

4. Date, C. J., Introduction to Data Base Systems, Addison Wesley, 6th Edition, 1992.

5. Elton, J. E. and Gruber, M. J., "Finance as a Dynamic Process" Prentice-Hall, Englewood Cliff, NJ, 1975.

6. Falconer, K. J., "Fractal Geometry: Mathematical Foundations and Applications", Wiley, New York, 1990.

7. Fama, E.F., Risk return and equilibrium, some clarifying comments, Journal of Finance 23(1), 1968, 29-40.

8. Freeman, J. J., "Principles of Noise", John Wiley & Sons, Inc., New York, 1958.

9. Gardner, M., White and brown music, fractal curves and 1/f noise. Scientific American, April 1978.

10. Gell-Mann, M. "The Quark and the Jaguar", W. H. Freeman & Co, New York, 1992.

11. Gell-Mann, M and Lloyd, S., Information measures, effective complexity and total information, Complexity 1(6), 1996, 44-62

12. Hutchinson, J., Fractals and self-similarity, Indiana University Journal of Mathematics, 30 1981, 713-747.

13 Lintner, J., The valuation of risk assets and selection of risky investments in capital budgets, Review of Economics and Statistics, 47(1), 1965, 13-37.

14. Lintner, J., Security prices, risk and maximal gains from diversification, J. of Finance 20(4), 1965, 587-615, 13-37

15. Mandelbrot, B.B. and Van Ness, J. W., Fractional Brownian motion, fractional noises and applications, SIAM Review 10(4), 1968, 422-437.

16. Mandelbrot, B. B., "The fractal geometry of nature", W.H. Freeman & Co, New York, 1982.

17. Mandelbrot, B.B., Self-affine fractals and the fractal dimension, Physica Scripta 32, 1985, 257-260.

18. Markowitz, H.M. Portfolio selection, Journal of Finance, 1(1) 1952, 77-91.

19. Myers, S. and Turnbull, S. M., Capital budgeting and the capital asset pricing model: Good news and bad news, Journal of Finance 32(2), 1977, 321-332.

20. Priestly, M. B., "Spectral Analysis and Time Series", Vol 1, Academic Press, New York, 1981.

21 Refines, A. N, and others, Neural methods in financial engineering, A study in Methodology, IEEE Trans on Neural Networks., 8(6), 1997.

22. Robinson, F. N. H., "Noise and Fluctuations", Clarendon Press Oxford, 1974.

23. Sharpe, W. F., Capital asset prices: a theory of market equilibrium under conditions of risk, Journal of Finance 19(3).

24. Sharpe, W.F. and Alexander, G. L., "Investments", 4th edition, Prentice-Hall, Englewood Cliffs, NJ, 1990.

25. Smith, Adam, "An Inquiry into the Nature and Causes of the Wealth of Nations.", 1776

26. Silberschatz, A, Peterson, J. L., "Operating Systems Concepts.", 5th Ed, Addison Wesley, Reading, Mass., 1998

27. Tobin, J., Liquidity preference as behaviour towards risk, Review of Economic Studies 26(1), 1958, 65-86.

28. Voss, F. V., Fractals in nature: From characterization to simulation, in "The Science of Fractal Images", H. Peitgen and D. Saupe Editors, Springer-Verlag, New York, 1988.

29. Voss, R. F., 1/f noise in music: Music from 1/f noise, J. Accoustical Society of America, 63, 1978, 258-263.

□