A Fundamental System Hypothesis relating Resources, Risk, Complexity and Expected output in Agent-directed Systems

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 agent-directed systems. The general veracity of this Hypothesis appears such that it could be considered a Fundamental Law of Systems. The risk measure is 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 scheduling & utilization complexity. Complexity is defined as a specification length after Gell-Mann. Both complexity parameters measure complexity in the system's environment-coping procedure that monitors an often close-to-random time function representing the unfolding environment. The Hypothesis is expressed as a mathematical relationship that reduces to numerical values for specific system circumstances. 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, fractal, Hurst constant, time function, random walk, resources, risk, spectral density.

1.0 Introduction

In recent decades the complexity of systems, particularly computer systems and systems containing major computer subsystems, has very greatly increased. At the same time complex systems are increasingly having to cope with serious 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 in bioengineering, and systems in critical manufacturing processes. With all of these systems, despite their complexity and associated risk, there are always directing human agents in the background. 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, complexity, risk and resources, which governs all systems. However, we are confident of the general veracity of this hypothesis, so that it could also be considered as a Fundamental Law of Systems. This Systems Hypothesis is of course relevant to computer systems and software engineering, and the author believes that persons familiar with its workings will be better equipped to think about, analyse, create and understand the implications of, and manage computer systems of all kinds.

There seems to be no record in the literature on systems, particularly computer systems, of work done to relate coping complexity to 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. Their results are widely used in the financial arena [3, 19, 21] and are standard material in undergraduate finance courses. 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. But most importantly, it brings complexity in a highly specific manner into the overall scheme of things along with system resources, risk and output, for systems in general.

2.0 A Fundamental Systems Hypothesis

At the outset, we consider a system to be an entity that functions under the direction of one or more agents. A system also employs resources in converting a set of inputs, informational or material or both, to a set of outputs. The system outputs are normally of greater value to the agents than the total inputs, so that is there is a net positive output value, measurable by the agent, associated with system operation. In addition a system is usually made up of a network of subsystems, each of which can be considered to be a system; all or part of the outputs from one or more subsystems can be the all or part of the inputs to another subsystem. With a subsystem the agent is not necessarily human.

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

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

* V is the dependent variable equal to the expected net value per unit time produced by the system.
* R is an independent variable equal to the value of the resources of the system directed towards exploiting an environment.
* r_P is an independent variable equal to a measure of the risk 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 to a different operating environment.
* r_N is an independent variable equal to a measure of the risk in the system's operating environment of loss of value, for risk it can not pay to run (negative risk) but can pay to prevent; r_N will frequently be zero; the independence of r_N is such that it can 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 frequently zero.
*c_m is an independent variable, the effective monitoring complexity, equal to a measure of monitoring or checking complexity, in one or more environment-coping procedures that are part of the system.
*c_m is an independent variable, the effective scheduling complexity, equal to a measure of resource scheduling & utilization complexity, in one or more environment-coping procedures that are part of the system.
Note: c_m and c_s can sometimes be combined into a composite effective coping complexity c_c, replacing $a_1c_m + a_2c_s$ in (1) by ac_c.
* k, a_1, b_1, a_2, b_2, and b_3 are constants, for a specific class of system and environment.

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

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

This means that for a system in a given environment with positive risk r_P, output V increases with increased resources R applied. Also, for a given R, if the agent shifts the system to an environment with higher r_P, output will increase.

(b) $V = K\,R\,(1 + a_1c_m + a_2c_s)$ (3)
or $V = K\,R\,(1 + ac_c)$ (3a)

This means that for system with a given environment-coping procedure (or procedures) with effective complexity c_c, output V increases with increased resources R applied. Also, for a given R, output will increase if the effective coping complexity c_c in the coping procedure is increased by the agent.

(c) $V = K_3\,(1 + b_1r_P - b_2r_N + b_3P + a_1c_m + a_2c_s)$ (4)
or frequently $V = K_3\,(1 + b_1r_P + ac_c)$ (4a)

This means that for a system with a given environment-coping procedure (or procedures) with effective complexity c_c, shifting the system into an environment with higher positive risk r_P will increase output V. For a system in a given environment with positive risk r_P, increasing the effective coping complexity of the environment coping procedure will increase output V.

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

The exact meaning of the resource measure R (and P) and net output measure V is important. The units of both R and V are the same, namely units of value. With respect to a human agent, the quantity Q of any commodity can serve as a measure of value, e.g. quantity Q salt, quantity Q of 64-Mbit DRAMs, by comparison to that quantity S of a commodity that serves as the standard of value, assuming that the quantity Q can be freely exchanged by system agents for a quantity of the standard. Thus the ultimate measure of R and V will be in units of a (standard) currency, although we can use any other units, such as quantity of any commodity, provided there exists some potential exchange rate (that is, the price), even if not currently known, between the commodity and the standard.

It should be clear that time must be considered as a resource when dealing with systems. All systems have a finite life, so that the resources R of the system are consumed over the system's life, this consumption affecting net output V. Suppose a system generates a specific net output v_1 in a time t_1. If system operation is changed so that it now takes time $2t_1$ to generate a net output v_1, resources $v_1/2$ have been diverted from the system over time period t_1. Thus time must be considered a resource.

Risk versus resources

This part of the Systems Hypothesis for the most part shows how output V depends on resources R applied and the positive risk r_P in the operating environment of the system. But before we proceed to analyse risk versus resources, we need a measure of risk.

Risk is normally run to secure some gain, and is therefore of primary interest. 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 suffer unpredictable losses 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 one is counterbalanced, at least partly, by a gain 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.

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.

(c) Neutralize the risk, by applying a complex environment coping procedure or procedures to detect either early hazard warnings, in time to undertake preventative action, or exploit compensating opportunities. For some risks this is not possible.

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. This can also be looked at statistically in terms of the frequency f of a hazard's occurrence over period of time T, together with size L of the loss. Thus over T, Lf is the total loss and thus a simple measure of the risk, that is, Lf/T, or Lp per unit of time. If the system is exposed to a set of hazards over T, each of probability p_1, p_2, ... and with associated losses L_1, L_2, ..., the aggregate risk is $L_1p_1 + L_2p_2 + ...$ Thus a simple measure of risk is the 2-tuple (p, L) or the product pL. Unfortunately, with complex systems these simple measures can be seen to be inadequate.

With any system, losses due to risk exposure show up as a reduction in net output V (but notice that since V is a net quantity, a V reduction may have its source in a loss in R). However remember that 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. This complicates matters very considerably. Suppose a situation with loss probabilities p_1, p_2, ... and loss sizes L_1, L_2, ... over a time T with a total loss 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 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

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

so that there is no net loss over time T, 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 - most will agree 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, involving many time units (e.g. years, minutes) the total net output of a system is W. Suppose that over the same time the system was exposed to unpredictable losses and gains, and 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 as part of the system resource R in the next time unit (that is, the equivalent of reinvestment in financial systems, or growth in economic or biological systems), then the average or expected net output V per time unit must be W/T. (If there was reinvestment/resource growth during T then V must be computed from W in a non linear fashion; but as this does not affect the essense of the argument and risk measure being developed, we ignore reinvestment/growth as an irrelevant complication.)

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

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

where $L_1, L_2, \dots$ are deviations downward (losses) from the average output V, and $G_1, G_2, \dots$ are deviations upward (gains) from V. These outputs can occur in some unpredictable order, and we will have

$j + i = n$, and

$(L_1 + L_2 + \dots L_i) = (G_1 + G_2 + \dots G_j)$

since we must have:

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

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 V in a given time unit (year, minute, etc), and it is such losses with respect to an average output 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 ($L_1, L_2, \dots G_1, G_2, \dots$) to give the Mean Deviation (MD) risk measure.

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 ($L_1, L_2, \dots G_1, G_2, \dots$) to give the Standard Deviation (SD) risk measure.

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.

To be quite specific about the meaning of these risk measures, suppose a 20 unit time period T, with a net output W = 2000 over T. Suppose also 5 time units each with a loss 15, 5 time units each with a loss 5; and furthermore suppose 5 time units each with a gain 5 and 5 time units each with a gain 15; all gains and losses are with respect to the expected output V and are assumed to occur in some random order.

The expected output V per time unit is 2000/20 or 100, and the average deviation is (15*5 + 5*5 + 15*5 + 15*5)/20 or 10, giving an MD-risk of 10. Thus in any time period looking forward, there is a 50% chance of a loss of 10, and 50% chance of a gain of 10; there is also a 25% chance of a loss <10 (i.e. 5) and a 25% chance of a loss >10 (i.e. 15), with respect to the expected V of 100, all of which make sense. (But notice, a person who considers only the losses and ignores the gains will find total losses over the 20 time units of 15*5 + 5*5 or 100, or an average of 5 per time unit, which will be half the MD-risk, and less than half the SD-risk.)

The standard deviation of the 20 deviations above is sqrt(5*225 + 5*25 + 5*25 + 5*225)/20) = sqrt(125) or 11.2, giving a SD-risk of 11.2. Thus there will be 50% chance of a loss, but 34% chance of a loss less than 11.2 and 16% chance of a loss greater than 11.2.

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.

The influence of resources on risk

We must now show using MD (or SD) risk measures r_P and r_N, that $V = K_1R(1 + a_1r_P + a_2r_N + b_3P$. Readers will be aware that 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

$$I = S(1 + i + kr) \quad \text{......} \ (5)$$

where i is the rate (expressed as a fraction of S) of return, or interest rate, in the absence of risk, and kr 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 general system positive risk r_P.

This relationship (5) is basic for financial systems. For example, a principal sum S invested in a long bond environment (higher fluctuation or SD-risk) will give a higher average income than S in medium-term bond environment (medium SD-risk), which income in turn will be higher than from P in very short term bond environment (little SD-risk). In general, the rule that return on investment of a given principal sum S increases with the associated 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 and the the (positive) risk of the operating environment, applicable to all systems, but true only for risk of a specific type. This 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.

Now consider two environments E_1, and E_2. E_1 is risk free and gives a fixed output F per time unit for application of a system with resources R which regularly transforms inputs N to outputs U. E_2 is the same as E_1 except that in E_2 the transformation of N to U is risky. Carrying out an operation in a risky environment must generate more value than carrying out exactly the same operation with the same resources in a risk-free environment. Hence the value output in E_2 must be higher than F as follows. Suppose gross output value in E_2 is F + G per time unit, in circumstance where the risk resulting from the environment E_2 is run but the hazard does not occur. Thus G is a gain in gross out 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. Either way, on average, there is a cost with respect to the extra gross output G.

Choice-1. Although it is possible to run a risk successfully during some executions of a system process, it is not possible over a significant time period. When a risk is run repeatedly, losses must occur over time, which affects system net output value. If the loss per time unit due to the hazard occurring is L_r, then the value output from running the risk in reality will on average be F + G - L_r per time unit. (Running the risk with insurance is essentially no different from running the risk; with insurance the inevitable long run losses are simply paid for in a regular way. Risk can also sometimes be reduced by counterbalancing it (classification A(c) earlier), but can rarely be eliminated; assume in this case of running the risk, that it is a fundamental risk that cannot be reduced further by counterbalancing.)

Choice-2. In order not to run the risk while preserving transformation of N to U and the extra 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 applying additional material resources to the hazard, which we assume costs L_p per time unit, in order to maintain transformation of N to U and achieve gross output F + G; the risk may be prevented by investing 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.

An example may clarify: Suppose a ferry system operating in risk-free environment E_1 that takes people from one side of a river (input N) to another side (output U) with constant value F generated per unit time. Suppose now the ferry operating environment shifts to environment E_2, similar to E_1, in which it continues to carry out the same operations of E1 to take people across the river; however suppose that in E_2 the river-crossings are carried out at a risky place and that the value output is now F + G per unit, where an additional output value G is generated per time unit for getting passengers from one side (system input N) to the other side (system output U) in the case where the risk is run but the hazard never occurs.

Choice-1. If passengers run the ferry risk, although sometime they will escape hazard and generate gross output F + G per time unit, on average there will be losses L_r per time unit, reducing the output to F + G - L_r per time unit

Choice-2. Passengers might avoid the crossing risk and take a long way round to a bridge at extra cost L_a per time unit, reducing output to F + G - L_a per time unit; alternatively the ferry risk might be preventable by directly applying material resources P to the ferry or crossing place, corresponding to a cost L_p per time unit, to make the crossing safe, reducing output to F + G - L_p per time unit.

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 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 more positive (a very important point), 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 (also a very important point), while maintaining the transformation of inputs to outputs, by either avoiding the risk or investing resources P to prevent the risk.

***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 run, due to incompleteness of agents' education.

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 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 by shifting R, from an environment E_1 with no risk and output F, to a similar one E_2 with output $F + G - L_r$ 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%. (Referring to the ferry example earlier, in E_s half the ferry runs per time unit are at a risky crossing, half are not). 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/2.

It is now important to grasp that, because of agent competition, in every similar but naturally occurring environment with risk $L_p/2$, the output for the system, with R applied, must still be G/2. We put this forward as a premise. From this it follows that output 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.

It should also be clear from a similar analysis of an environment E_3 containing only a risk it can not pay to run but pay to avoid or prevent, that we must have

- $V = KR(1 + b_1r_P - b_2r_N)$ (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.

Furthermore, since it some environments it will be possible to apply additional resources P to prevent (or reduce L_r) without affecting G, with the amount of L_r offset being proportional to P. Hence V must rise as P increases, so that the full relationship must be

$$V = KR(1 + b_1r_P - b_2r_N + b_3P) \qquad (2)$$

Expression (2) states that as the resources R of a system are applied to riskier and riskier environments up to limiting environment E_m, with maximum risk r_m, output increases. However with environments after environment E_m, if the increasing risk is due to addition of negative risk to r_m, output will peak with environment E_m and then fall off. 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".

However expression (2) has more to say. For any environment with either positive risk, or negative risk, or both, it may be possible to increase the return by applying resources P to the direct reduction of the (losses in output due to the) risk, whether positive risk or negative risk. Where the environment does not allow this, P will be zero.

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 will 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 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 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 output of 250 and an MD-risk r of 100. The output 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 if the hazard does not occur. But will 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 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:

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

if the risk variable r_P is a mean deviation measure (MD-risk) expressed in per unit values. (MD-risk of 100 is 0.1 per unit of 1000.)

[For readers familiar with financial analysis, the following example using MD and SD-risk and risk it can pay to run repeatedly may clarify better. Suppose E_1 involves \$1000 in risk free T-bills at 5% p.a. and E_2 \$1000 in long bonds at 8% p.a. or nominal income of \$80, with long bond annual return or income fluctuating (20% of \$1000) between \$280 and -\$120, giving an MD-risk of \$200 annually. Hence F is \$50 or 5%, and if the hazard (annual return of -\$120) never happens, average extra return for E_2 over E_1 is 30 + 200/2 or \$130 or G per annum. L_r is the average loss in E2 per annum, equal to half the MD-risk of \$200. So $L_r = 100$, and $G - L_r = 130 - 100 = \$30$ per annum or the excess over 5% return or \$50 income (F) per annum due to risk in E_2. Since G-L_r is positive, this indicates bonds are a risk it can pay to run repeatedly if T-bills return 5% in E_1. The expected income in E_2 will be $F + G - L_r = 50 + 30 = \$80$ per annum, or expected return of 8%. SD-risk is also \$200 annually in E_2. Using MD-risk for r_P expression (2a) expressed as per unit values of R (MD-risk of 200 is 0.2 of 1000), the relevant version of expression (2a) will be

$$V = 0.05\, R\, (1 + 3.0\, r_P) \qquad (2d)$$

which will be applicable to other similar environments, e.g. medium term bonds, or synthetic environments constructed by placing part of R in (say) long bonds and part in T-bills [3, 24]. This illustrates that expression 2a can be put to real practical use.]

3. Resources versus Monitoring & Scheduling Complexity

This part of the Fundamental Systems Hypothesis states that for a system with a given environment coping procedure with effective complexity c_c, and given resources R, output V can be increased by increasing the effective complexity c_c in the coping procedure in accordance with

$$V = KR(1 + a_1c_m + a_2c_s) \qquad (3)$$

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

To show that (3) holds generally we need to solve two problems. First we need a reasonable and sensible measure of the complexity of (or in) a coping procedure. Second, for a given level of system output we need to show that increasing coping procedure complexity will increase output.

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. A procedure may be said to cope with its environment 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 (some of 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. In addition, resource scheduling & utilization complexity in the coping procedure that enables system resources to be better scheduled and utilized over opportunities and pitfalls in the environment will enable increased environment exploitation per unit time and so increase output.

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 we have a financial management system, the data stream is the prices coming from the markets, if it is the paging supervisor of a virtual operating system, the data stream consists of page access data, if is a mobile robot systems, the data stream consists of image data, and so on. Data that has already arrived constitutes a history of past environment unfolding, i.e. the historical data string.

We can characterizes 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 important 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 at some point 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 that if the effective complexity of the incoming data stream is c, then the complexity in the coping procedure needed to monitor the regularities in the string must also be c or proportional to it.

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 prevent risk of 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.

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:

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

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

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

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:

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

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 zero. Mathematically, a time function S(t) with H not equal to 0.5 is a generalization of a random walk, often referred to a 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 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 an 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 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 $I_{i1}, I_{i2},\ldots I_{iT}$ 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 $\langle(W_i)^2\rangle$ or:

$$\langle I_{i1}^2 + I_{i2}^2 + \ldots I_{iT}^2 \rangle$$
$$+ \langle I_{i1}I_{i2} + I_{i1}I_{i3} \ldots + I_{i2}I_{i1} + I_{i2}I_{i3} \ldots\rangle$$

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

$$\langle I_{i1}^2\rangle + \ldots \langle I_{iT}^2\rangle$$
$$= T\langle I_{i1}^2\rangle \quad = TK^2$$

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

$$\langle (I_{i1}I_{i2} + I_{i1}I_{i3} \ldots + I_{i2}I_{i1} + I_{i2}I_{i3} \ldots)\rangle$$

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

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

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

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

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

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

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

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

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

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:

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

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 can 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 time functions, 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 it seems unlikely that any functions encountered in a system's environment will have H exceeding 1.0 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, 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.

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 imperatives of the form:

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

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

...

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

The coping procedure may be said to be saturated with monitoring complexity if all regularity inherent in the time function being monitored is listed in the procedure's order-specifications. For a saturated procedure, it follows that the sum of the lengths of S_{A1}, S_{iA2}, S_{B1}, S_{B2}, ...] will be equal to the effective complexity c of the time function, since these specifications reflect all the regularities in the function. If the procedure is unsaturated, we define the sum of the lengths of S_{A1}, S_{A2}, S_{B1}, S_{iB2}, ... as the effective monitoring complexity c_m of the procedure. Adding additional conditional-imperative pairs to the coping procedure, with order specifications taken from the function, will then increase the procedure's effective monitoring complexity up to its maximum of c.

The total additional expected output E of a system with this procedure, as a result of applying the procedure must then be:

$$E = [p_A G_A - (1-p_A)L_A] + [p_B G_B - (1-p_B)L_B] + \ldots$$

when the coping procedure is saturated. When it is unsaturated with effective monitoring complexity $c_m < c$, it is reasonable to average the additional expected output as $E = c_m(E/c) = Ka_1c_m$, where K and a_1 are constants. Thus the output of the system will increase with the effective monitoring complexity of a coping procedure adapted to and monitoring the unfolding environment. However since the system will have resources able to generate output without the application of the coping procedure, as covered in Section 2, additional expected output in the presence of the coping procedure must depend on

$$KR(1+ a_1c_m)$$

where the procedure contains monitoring complexity specifications measured by c_m.

Effective resource scheduling & utilization complexity of a coping procedure

However relating expected output to effective monitoring complexity c_m in the coping procedure is not the end of the matter. We should also take into account the complexity of any resource scheduling & utilization (RSU) specifications embedded within the coping procedure for efficient scheduling and utilization of system resources that carry out operations of types A, B, C etc. However, the complexity, inherent in scheduling & utilizing system resources for the operations, has nothing to do in general with the unfolding time function being monitored. The time function merely signals each of opportunities/pitfalls specified by S_{A1}, S_{A2}, S_{B1}, etc. to the coping procedure. In some cases the time function may be so ordered, that is, effectively non-existent, that c_m is zero. Specifications for efficiently scheduling and utilizing system resources for the operations carried out will be implemented in the coping procedure in the instructions for

<exec operation-type-A>, <close operation-type-A>
<exec operation-type-B>, <close operation-type-B>
...

Let the total length of the RSU specifications in the coping procedure be c_s, the effective complexity of the RSU specifications in the coping procedure.

When a specific opportunity/pitfall specified by (S_{K1}, S_{K2}) is signaled, even where the monitoring complexity c_m is very low or zero, the coping procedure must allocate and schedule resources to exploit/avoid it. We can assume that a resource designed to handle a given opportunity/pitfall can normally do so with zero RSU complexity in the coping procedure, typically no more than Start/Stop commands. However, for a given monitoring complexity c_m, that is, for a given set of opportunity/pitfalls specified by (S_{A1}, S_{A2}), (S_{B1}, S_{B2}) etc., that the coping procedure is in a position to detect, resources available for allocation by the coping procedure will often be insufficient.

If the resources can handle, with zero scheduling complexity, only a portion of the opportunity/pitfalls detected by the coping procedure, and if the percentage utilization of the resources in handling this portion is less than 100%, which is to be expected, then it will be possible for these resources to handle additional opportunity/pitfalls if the resources are shared between the opportunity/pitfalls, that is, if we increase the RSU complexity in the coping procedure for each of the operations A,B,C etc.

It follows that in general, when resources are held constant, expected output will rise with increasing effective RSU complexity in the coping-procedure, and decreased resources can be compensated by an increase in effective RSU complexity.

Hence the expected output must depend on the level of resources R, the effective monitoring complexity c_m and the effective resource scheduling & utilization complexity c_s in accordance with

$$V = KR(1 + a_1c_m + a_2c_S) \qquad (3)$$

In a given situation, either monitoring complexity may dominate (large a_1, negligible a_2), or RSU complexity may dominate (negligible a_1, large a_2), or they may both be significant. In many cases where c_m is zero, only c_s will matter. However it will often be convenient, at least where c_m and c_s do not matter much, to lump the two complexities together and rewrite (3) as

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

where c_c is a composite complexity, which we might call the effective coping complexity of the coping procedure.

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

Since we have now shown that

$$V = KR\,(1 + b_1r_P - b_2r_N + b_3P) \qquad (2)$$
$$V = KR(1 + a_1c_m + a_2c_s). \qquad (3)$$

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

$$V = K\,R\,(1 + b_1r_P - b_2r_N + b_3P)(1 + a_1c_m + a_2c_s) \qquad (6)$$

or

$$V = K\,R\,(1 + b_1r_P - b_2r_N + b_3P + a_1c_m + a_2c_s) \qquad (1)$$

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

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

output V is forced negative. In the analysis in Section 2 it was noted that a risk could be prevented by a 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 complexity c_m or c_s, V will stay negative. But it is known that risk can be mitigated by using a complex procedure, so that c_m and c_s ought to behave in the expression like P, as far as mitigating 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 is given by

$$V = KR\,(1 + b_1 r_P + ac_c) \qquad (4a)$$

if we use composite coping complexity c_c, assume no negative risk, and no resources applied directly to preventing risk. The full expression relating risk and complexity for a system with fixed (capacity) resources is of course:

$$V = K\,R(1 + b_1 r_P - b_2 r_N + b_3 P + a_1 c_m + a_2 c_s) \qquad (4)$$

6. Concluding Remarks

There appears to be some misunderstanding about such concepts as risk and complexity, so it is important to summarize clearly and exactly what the Fundamental System Hypothesis in expression (1) is saying. 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 to prevent).

First of all, for the system in any environment E_n, output 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.

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. If shifted to an environment with comparable positive risk r_P, but much increased 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 or both, losses due to the hazards associated with these risks actually taking place can be reduced by the agent without affecting 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 V will increase with increased applied risk-preventing resources P. P will normally be obtained from outside the system and will not affect R. Such resources P must be considered as a special category of resources separate from R. Nevertheless, total resources utilized by the system will of course be P + R. However, it may happen, when the agent realizes unexpectedly that the system is in an environment with large negative risk, that the agent cannot shift the system out of that environment, and that there are no resources available from outside the system to increase P to prevent the large negative risk r_N that will likely be making (expected) output V negative; in such cases the only option will be for the agent to divert (cannibalise) resources from R (reducing output) to increase P (increasing output much more). [A trivial example is the out-of-fuel ship in a remote isolated region burning its superstructure for fuel.]

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 (the unfolding time function is not a random walk.) A simple financial example would be funds placed in long bonds. The output would be improved by a monitoring procedure that could signal the market swings in advance. In practice bond fluctuations are close to a random walk so that this would be difficult, if not impossible, to do.

Finally, it says that for system with a given environment E_n with fixed risks, either the usual positive risks r_P, but also negative risks r_N, or with zero risk, for a given R, and for a coping procedure where monitoring complexity c_m is zero (because the environment is highly ordered and predictable) output can be increased if the agent increases the effective resource scheduling & utilization complexity c_s of the coping procedure, for example according to :

$$V = K\,R(1 + a_2c_s) \qquad (7)$$

This phenomenon is easily observed in computer systems, where it is common practice to increase system throughput by increasing effective RSU complexity. Obvious examples following expression (7) are multiprogramming operating systems [26] and file and database systems [1]. [A classic everyday example is the couple with extensive but different daily business and social engagements but only one car (R); they need a complex everyday-executed procedure for resource allocation and scheduling (high c_s) in order to manage (achieve acceptable V); acquiring a second car (much increased R) will instantly remove the need for the complex procedure (zero c_s) and fulfill the same needs (same V).]

For specific classes of environments, the constants and variable parameters in expression (1, 4) can be reduced to numbers and measurable quantities. Thus the Fundamental Systems Hypothesis can be subject to experimental verification. Given the manifold nature of the Hypothesis, many experiments will be required. 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. If there is a weakness, it is likely to do with the measures of complexity c_m and c_s; it is possible that future experimental work may require some minor modifications here.

References

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

2. Beaumont, G. P., "Probability and Random Variables", EllisHorwood 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 Cliff, 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.