COST BENEFIT ANALYSIS OF INFORMATION INTEGRITY

(Research Paper)

Vijay V. Mandke
Center for Information Integrity Research, India
vmandke@unitechsys.com

Madhavan K. Nayar
Unitech Systems, Inc., USA
mnayar@unitechsys.com

Abstract: The paper studies a specific research query of developing an analytical method for comparing two integrity mechanisms. Towards this a generic business process is modeled as an integral to a closed loop information and control system. Competitive advantage requires informational work from this business process information system (*IS*) is maximized. This *IS* is a multiple stage decision process and involves at each stage information origination and processing activities that are impacted by system environmental factors. This makes the business process *IS* view a continuous individual information originating and processing situation characterized by uncertainty and hence information errors leading to loss of Information Integrity at each stage. This is a structural variant of a traditional collective decision process based view of *IS*. For maximization of informational work (i.e., *use*) and for integrity analysis, how is one then to model this *IS*? In response to this query the paper develops the information Usefulness-Usability-Integrity paradigm, which offers determinants of information value. Recognizing that information origination is a costly activity, the paper then suggests in the form of cost-benefit analysis of Information Integrity a methodology to compare two integrity mechanisms. This is followed by development of equations for calculation of value of information and of improvement in value of Information Integrity due to additional information.

Key Words: Open and Closed Systems, Informational Work, Decision Making, Information Origination, Uncertainty, Information Errors, Information Integrity, Usefulness-Usability-Integrity paradigm, Information Integrity Attributes, Information Integrity Risk, Value of Information, Improvement in Value of Information Integrity

1. INTRODUCTION

Business organizations and their systems, sub-systems, and their components are getting recognized to be "open systems". Unlike closed systems, open systems have purpose (objective), possess porous boundaries with their environment, and process information [9]. With system integration, this makes for business activity *emphasizing* information and comprising informational and physical work systems. For competitive advantage, requirement is to maximize informational work (IW) comprising activities of: (a) generating from business process activities raw data/information in a complex and changing real world environment characterized by uncertainty and hence errors, and (b) processing this information on current basis for undertaking planning and evaluation of business process design alternatives and delivering selected information decision for control implementation at the physical work system.

2. BUSINESS PROCESS *IS* VIEW– INDIVIDUAL INFORMATION ORIGINATING AND PROCESSING SITUATION

The research presentation on "Information Envelope and its Information Integrity Implications" at IQ2001 [9] studied modeling a generic business process as an integral to a closed loop information and control system constituting a business process *IS* view. Most information processing involves some type of data conversion to information in *use* and, therefore, is closely related to a decision process with an objective. Even when the information is transmitted without changing form, as in a communication system, the issue is to decide the purpose or objective of the transmission [9, 10].

Traditionally, decision process is viewed to comprise stages of forecasting, evaluation of alternatives and selection [10]; information being considered basically as function of "source" (i.e. as "data") and at the most of "source" and "process". However, business process *IS* view is an open system; information being function of "source", "process" and "recipient" (i.e. customer). For it more workable model of a decision process spans multiple stages. They are: based on long term goal set, *obtaining* 'many factors' & 'multiple criteria' characterizing problem (task) complexity; from multiple criteria, *recognizing* (deciding) on operable goal; from operable goal statement, *defining* planning & design constraints and opportunity spaces; from 'many factor' information variables characterizing problem complexity, *culling out* useful (relevant) information variables; *recognizing* relationships (interdependencies) between culled out information variables; *developing* state transition models defining dynamic behavior of culled out state (information) variables; and undertaking customized planning & design for *generating* alternatives for evaluation and final selection of flexible information decision for control implementation [9].

What is significant is this multistage *IS* view, among other stages, involves identifying operable goal, originating information, and generating alternatives, all, "endogenous" to *the* decision situation - a structural variant from traditional view of decision process, which (in the manner of a closed system) is concerned only with alternatives and information that are "already" generated exogenous to the decision situation and hence is a "collective" decision process. Thus, in multiple stage decision process based business *IS*, what we have before us, is an *individual information originating and processing situation*.

3. BUSINESS PROCESS *IS* VIEW – CONTINUOUS INDIVIDUAL INFORMATION ORIGINATING AND PROCESSING SITUATION IN THE PRESENCE OF UNCERTAINTY

As with traditional *IS* which is a "collective" decision process, this multiple stage decision process *IS* is also characterized by uncertainties due to system environmental factors of 5"C"s, namely, complexity, change, communication, conversion, and corruption [7, 9]. Specifically, acting externally and internally, 5"C"s introduce in *IS* uncertainties observed traditionally and beyond. Thus, at the physical operations stage and at the physical variable control stage, uncertainties are introduced due to input noise, process parametric noise, and measurement noise. At all the control levels (physical variable, transaction processing and management decision controls - which in the wake of "application" emphasis are so impacted), uncertainties are due to information overload, lack of standardization, lack of relationship in data in several applications, errors in hardware, software errors, data entry errors, or accidental or intentional failures, etc. Further, uncertainties are also introduced due to incomplete knowledge of system dynamics and due to judgmental errors both at managerial decision level controls comprising human-machine interfaces and at higher level controls like production control, which apart from human-machine interfaces, in all probability, may even include humans as part of the process to be controlled.

And then there are the multiple individual decision process stages, which are also impacted by 5"C"s. Briefly, the business process *IS* view is a complex *IS*. This complexity introduces at the operational level, hitherto unknown complex error mechanisms coming from system development and implementation life cycle phases and that, too, coming with delay. At the stages of operational level and at control levels, it introduces errors that arise due to failures of embedded systems; and, at all business *IS* stages, errors due to emphasis on system integration maximization. The later category of errors is on count of resulting system interfaces exposing innermost system modules to uncertainties due to external system environmental factors and vice versa. Further, particularly at the decision process stages, the complexity introduces errors due to information processing under: multiple goals (implicit goals included); multiple factors & multiple criteria (goal descriptions); a large number of interdependent information variables, varying with time and not completely and correctly observable; and system dynamics not well understood (reality is not passive but – to some extent - active). All this results in information errors leading to loss of Information Integrity in business process *IS* view and in information therefrom [9].

For the purpose of present investigation, this draws attention to an important question: what if, in above *IS*, the "goal" leading to usefulness factor with reference to information originated, though given, continuously needs adjustment due to constantly changing environment (as very well can be the situation in say a service sector – e.g., medical treatment of an adult patient) or is not known or is out of date or is by itself complex? All these are the conditions to be observed in the real world problem solving. To further emphasize, even from a conservative point of view a large, semantically complex, time-pressured, tightly coupled, high consequence, high-reliability engineering system, in the wake of unclear goal statement (implicit goals inclusive), is observed to run a risk, in the fashion of an open system, of taking a life of it's own [5, 9]. In such case then the tasks of culling out the relevant facts (usefulness factor with reference to data and information variables) and of defining their interrelationships under the subsequent decision stages of the business process *IS* view cannot be treated as static ones determined uniquely and exogenously as in case of closed systems, but would acquire dynamic - open and *endogenous* to the decision making situation in that – character in the presence of 5"C"s, and they (data and information variables) would need to be *continuously* originated and processed. This reality leads to model information processing under the business process *IS* view as a *continuous individual information originating and processing situation in the presence of uncertainty,* so as to account for demands of continuously determined specific goal based individual situation in a complex and changing environment.

4. SIGNIFICANCE OF EFFICIENT AND ECONOMIC PROCESSING OF INFORMATION: ON CRITICALITY OF INFORMATION INTEGRITY FOR COMPETITIVE ADVANTAGE

For competitive business advantage, this *IS*, ridden with uncertainty and errors, must process information efficiently and economically. This makes Information Integrity a critical *IS* factor. To elaborate, consider any *IS*, say, an educational system and its sub-system, namely, an examination system. A common "measure" for a learner (candidate) performance is % marks providing a basis for *comparing* two or more candidates for their academic standing (specialization) so as to facilitate the candidate *selection* decision. In order to function *easily* in the recruitment market, the recruiter (employer, i.e., the customer of educational institution products) requires a common denominator to work with: % marks obtained by the candidate. Such information is, indeed, expressed in a form that makes it particularly *usable* in the context of "candidate performance ranking mechanism for comparison and selection decision." Thus, what is important is, the success of information "I" (% marks) as a medium of exchange and unit of measurement (measure) *is* in the fact that *it (information "I") minimizes transaction or comparing and selection costs. This* efficiency and economy in the market information gathering (originating) and processing systems that then provides the engine for both internal and external performance measuring systems based on concerned information "I" getting emphasized as the market grows in complexity. This explains the bottleneck character of the information used for the purpose.

There is little more to the analysis being pursued. Increased importance of examination information is also accompanied by useful role this information plays in allocating resources (e.g., improved T-L production factors, improved theory and practice of examinations, etc.) at two levels. Specifically, efficient resource allocations in any product or factor market *require* competition. Competition requires, as necessary conditions at least two things: (a) information about market imbalances (examination results' imbalances for different subjects, programs and educational institutions in this case) indicating improvement (business) opportunities, and (b) information on working mechanisms (knowledge capital) for implementing improved T-L production factors and for improved implementation of examinations, to exploit opportunities. For efficient processing of information, trade off *has* to be between costs associated with originating and processing of information and loss due to incorrect information, in that, the *IS* which, for a certain kind of information origination, processing, storage, distribution and discard, is able to arrange them (costs) at the lower level *will* tend to prevail. In view of this it follows that, *to compete successfully*, the information regarding (a) measure of candidate's performance — the aggregate, (b) examination results' imbalances — the opportunities, and (c) knowledge of working mechanisms for resource allocation, i. e., the knowledge capital *must* have "integrity". In other words, each detail in each of these information statements (the measure, the opportunities and the knowledge capital), and not only the bottom line statements, must be accurate, consistent and reliable; as it is only through ensuing of optimal integrity that it is possible to achieve efficient and *economic* processing of information (in respect of examination results) in the recruitment market described above.

The case of a recruitment market, candidate performance information statement in terms of % - marks-based examination measure, recruiter as its recipient, and education system comprising competing educational institutions, programs and subjects as analyzed here is *only* illustrative and incidental. The central point is if *IS* is to originate and process information efficiently and *economically*, that is if the decision process - that the *IS* is - is to deliver information decision as output *so that* recipient achieves maximum information *use* for the decision objective at hand (and this *IS* must do if it were to be competitive), then it is *fundamental* that the information, which constitutes the bottleneck resource, *is* with integrity. In other words, dependability and trustworthiness, i. e., Information Integrity is the critical factor in controlling the quantum and economy of originating and processing information for *use*, i. e., for strategic and competitive advantage in complex and changing business environment.

5. INFORMATION INTEGRITY – EXISTING PERCEPTIONS: THE MAIN LIMITATIONS OF

Literature reports integrity studies from different angles: security based definitional approach to integrity, auditing research, process centered quality approach, noise reduction based technology under communication theory, and the Savage (Subjective) Expected Utility (SEU) theory under decision-making. Briefly, in computer science security is taken to mean confidentiality, integrity and availability with the word "integrity" describing a range of requirements [2, 7]. Further, database integrity models and methods, while context specific, do not lend themselves to any comparative, analytical studies [14]. In accounting/auditing research, with respect to accounting information, relevant part of the internal control structure is made up of three basically ad hoc categorizations: the control environment, the accounting system, and the control procedures. This offers a way of structuring the analysis of different possible control mechanisms but with no explicit coupling to cost and benefits in the sense items in different categories can be compared [11]. Same limitation is with the qualitative COSO report that sees internal control, from the management angle, as consisting of five interlocking factors: monitoring, information and communication, control activities, risk assessment, and control environment [1, 14].

Quality paradigm has two aspects; namely, (a) quality assurance concentrating on the process and attempts to ensure that it is done correctly, and (b) quality control to ensure that the product delivered to customer (recipient) is correct, where the term 'product' represents a system or component or service. In practice, however, the quality paradigm operates in the 'standard' product mould, emphasizing incremental changes, and sees its operable goal as 'reduced defects'; thereby emphasizing cost reduction aspect but not the cost-benefit angle. This leaves the quality emphasis weighing more on the side of 'process'-centered issues rather than 'product'-centered issues [13]. In the wake of ever-present channel noise, *IS* model in communication theory is concerned with the problem of reproducing at one point (destination) either *exactly* or approximately a message selected at another point (source). The noise reduction technology envisaged is optimization of variable parts of *IS* (encoders and decoders) so as to improve reliability, increase the data rate, or decrease the cost. This *IS* model does not take a decision process based view of the message through the channel. In fact, although the measurements in information theory are significant to communications engineer, they are not related to decision issues, except by chance [4, 14]. Accordingly, then there is no reference to the cost-benefit framework for the degree of "exactness" of message achieved.

The SEU Theory analytically studies a decision process model under uncertainty based on the concept of information value and hence in the first instance seems to be an attractive proposition to study integrity issues. However, SEU maximization is descriptively *invalid* – falsified – as a model of how individual decision makers behave. Nevertheless it is descriptively *valid*, or at least constitutes the best alternative currently available, as a model of individual decision making when building theories of collective decision making at the market level. When dealing with a problem as at hand, this apparent paradox causes confusion [6]. Further, SEU Theory defines the monetary value of perfect information as amount of money which renders the decision maker indifferent between using and not using information; and thus does not consider in its treatment any explicit coupling to cost-benefit analysis for the information value it measures.

In summary, the *IS* models presently in vogue in integrity research literature do not account for the requirement of continuous origination of information endogenous to the specific decision situation. Their main concern is *only* that information technology accesses, communicates, processes and distributes the *already* generated information. With information technology costs ever decreasing, the information processing for decision-making is, therefore, taken as a *costless* activity; resulting in *IS* models having no explicit reference to cost-benefit of information processed. The reality and its requirements though are different. The *IS* under consideration comprises individual decision process stages that are characterized by activities of information "origination", which is a *costly* activity. This calls for cost-benefit analysis framework for information originated and processed, so as to work towards ensuring economic processing of information. As argued in Section (4), it is through control of Information Integrity that this economy is ensured; thereby the cost-benefit analysis framework required in fact being that of Information Integrity. Towards this objective the paper then considers development of information Usefulness-Usability-Integrity paradigm with the objective of describing the attributes of Information Integrity.

6. USEFULNESS-USABILITY-INTEGRITY PARADIGM

From 'many factor' information variables characterizing problem complexity, the business *IS* view under consideration has a requirement to cull out "useful" (relevant) information variables (Section (2)). Further, for competitive advantage, information is expressed in a form that makes it particularly "usable" in the context of mechanism for comparison and selection decision (Section (4)). In search of a structure for integrity objective, this provides a basis for the Usefulness–Usability–Integrity (UUI) paradigm. Specifically, usefulness refers to the *relevance* of the information for its intended purpose. For example, the recent history of a stock's price may be useful in deciding whether to buy or sell a stock. However, the recent history of the price of corn or oil may not be useful at all in deciding whether to buy or sell the stock. Against this, usability refers to *feasibility* factors such as availability, accessibility and understandability, which help make it possible and easy to use the information. For example, information may be usable because it is available on the Internet, because it is presented in an intuitively obvious format or because it can easily be imported into a spreadsheet or database.

Literature identifies a universe of information attributes; namely, accuracy, usability, reliability, independence, timeliness, precision, completeness, relevance, sufficiency, ease of understanding, freedom from bias, consistency, trustworthy, brief, etc. [3]. Appropriate attributes from these concerning context, goal, and nature of information use, i.e. relevance and feasibility of use, then can be categorized under the usefulness and the usability objectives. This facilitates a workable framework for defining intrinsic integrity objective in the form of accuracy, consistency and reliability attributes of information covering correctness aspect [7, 8]. Information requirements of usefulness, usability, and integrity are, then, the determinants of information value. Seen more critically, usefulness and usability factors are also defined by their respective information requirements. It goes without saying these information requirements must also have integrity. In other words, integrity attributes of accuracy, consistency and reliability are fundamental or basic to the information requirements of usefulness and usability and, therefore, to the value of information; and as a result a critical requirement of an *IS*. Further, as information value can be seen to define information use (IU) quantum, within above framework the integrity objective then can be seen as to optimize "IU" quantum for a given information processing situation, so as to offer competitive advantage.

7. DEFINING INFORMATION INTEGRITY ATTRIBUTES

At this stage, it should be of help to get a clearer view of integrity attributes of accuracy, consistency and reliability so as to be in a position to develop the equation for cost-benefit analysis of Information Integrity. Contrary to information "exactness" requirement as pursued by traditional Integrity research (Section (5)), the search is for "*correctness*" aspect of information requirement. For describing Information Integrity attributes, this suggests it to be useful to start by defining "error". What can be construed as an error? From the viewpoint of an external observer, an error can be seen as a failure to ensure an optimum, desired, or intended value (for a view, format, variable, or process, etc. as the case may be) that is correct given the circumstances (situation), the cause and form of error not withstanding. An error can occur only if there is an appropriate identified source of value (standard) to ensure on the basis of a documented state of events.

Within above framework then <u>Accuracy</u> attribute (A) is defined as the degree of agreement between a particular value and an identified source. It can be assessed by identifying the relevant established source (standard) and by determining an acceptable tolerance. Specifically, the identified source provides the correct value – preferably the value corresponding to the optimum Integrity (see Section (8)). Against this, <u>Consistency</u> (C) is defined as the degree to which multiple instances of a value satisfy a set of constraints. The multiple instances may exist across space (such as databases or systems) or over time. Consistency is then with respect to a set of constraints and data/information is said to be consistent with respect to a set of constraints if it satisfies all constraints of the data/information model.

Reliability attribute (R) is a little complex attribute to define. Traditionally, it is a large concern in system development lifecycle model and refers to a wide range of issues relating to the design of large systems (complex computerized information system (*CIS*) included), which are required to work well for specified periods of time. From this point of view for an *IS* the definition of reliability given as "accuracy with which information obtained represents data item in whatever respect the information system processed it" can be seen to define the reliability requirement for the *IS* as a whole [7, 8]. Then, as mentioned under Section (5), reliability is also seen as 'completeness' issue. Of course, the completeness requirement itself has two different aspects. One is that of "exactness" requirement. This requirement occurring due to the ever-present system "noise" is the main concern in communication theory and in security research as also in the "standard" product in high volume seeking business models under quality paradigm emphasizing "reduced defects" in system processing.

There is another aspect of "completeness" requirement, though. In the form of "observability", it is to be found in system theory. Specifically, the problem considered is that of state variables derived based on measured system outputs at several times and the knowledge of the system-forcing function (control) effort. It is conceivable that the structure of the system and/or measurements taken is such that the measurements do not contain *all* the information about the system states. The usual technique in systems engineering is to generate control efforts (strategies) based on measurements of system outputs. If the measurements are *missing* basic information on actual system response (that is, if there is information distortion), erroneous control efforts could be generated, which is not desirable; just as, if, in the *IS*, value of information element is missing from the information record, the desired information use (IU) value is not achievable, however high may be the information usability factor.

In other words, when concerned with reliability factor under *correctness* requirement of information, there are incompleteness issues due to "noise" and "distortion". For the purpose of the investigation at hand, whether "inexactness" due to the 'noise' factor or "incorrectness" due to 'distortion' factor, both result in information item exhibiting error and therefore loss of integrity. As a result, reliability attribute of "*correctness*" aspect of information requirement in considering 'completeness' must account for both these possibilities. It is within this framework then the reliability (R) can be heuristically defined as follows: Reliability (R) refers to completeness, currency and auditability of data/information. Specifically, data/information is complete when all component elements are present (effects both of distortion and noise are counted). Information is current when it represents the most recent value. And, information is auditable if there is a record of how it was derived and that record allows one to trace information back to its source.

For the analytical convenience, let us denote the reliability attribute defined based on "completeness" perception, which accounts for 'distortion' as mentioned above, by "R1". Further, let us denote the reliability attribute, which is with reference to "noise" factor and is based on "exactness" perception, by "R2". Then, the reliability attribute "R" should cover both "R1" and "R2" and is given by Equation (1).

$$\text{Reliability attribute index} = R = R1 \times R2 \qquad \text{.. Equation (1)}$$

What are defined are attributes of Information Integrity (I*I) for information value, i.e., for the content of information processed by *IS*. Content of information would, therefore, have I*I value as in Equation (2).

$$\text{Information Integrity} = I{*}I = A \times C \times R1 \times R2 \qquad \text{…….. Equation (2)}$$

With determination of I*I attributes and I*I, one can now proceed with cost-benefit analysis of I*I.

8. COST BENEFIT ANALYSIS OF INFORMATION INTEGRITY

Consider any information originating and processing stage (S_i) of the *IS* view as in Section (2). In its representation this *IS* can be seen as comprising a number of core *IS* models having data origin stage, data transformation/ conversion/ processing stage, pre- and post- processing communication channels (comprising medium and people), and output (i.e., data product, that is, information presentation, obtaining and *use*) stage. These core *IS* models may be repeated, paralleled, and interrelated. Output from one core *IS* model may become input to another [7, 10]. It is recalled that core *IS* model to which data and information are integral is modeled as a decision process (see Section (2)). To outline the cost-benefit analysis methodology of Information Integrity, one can consider such decision process. The decision purpose can be taken to process/transform/convert data as in core *IS* to deliver information decision (by itself an information) so as to achieve better information use (for example better control for improved customer service). Thus the purpose of processing data/information through the core *IS* can be taken as "improvement in information *use*", which in turn gives the strategic or competitive advantage.

It is understood that this "improvement" as a variable will be a function of the information (I) being processed under the stage $\{S_i\}$ and, accordingly, it can be represented by $[\Delta IU(I)]$. Let IUUB(I) denote the variable giving the upper bound of information use as function of "I" (given that such function can be defined). Let "$\alpha(I)$" denote usefulness factor and "$\beta(I)$" usability factor. Both factors, functions of "I", may take values between (0,1] and, accordingly, can be seen as appropriately defined proportionality variables. Then, the improvement in information use at stage (S_i) is given by Equation (3).

$$\Delta IU(I)\big|_{Si} = [\alpha(I) \times \beta(I) \times IUUB(I)]\,\big|_{Si} \qquad \text{............Equation (3)}$$

But, reality is different as core *IS* models are complex, open and impacted by 5"C"s and they *have* errors. As a result there *is* a question about the integrity of information "I". Specifically, suppose question is about the accuracy of information, and let "A(I)" denote the concerned integrity quotient, which takes values between (0,1]. Then, the gross "benefit" or improvement in information use from information processing at stage (S_i) would get modified to as in Equation (4).

$$\Delta IU\,(I)\big|_{Si} = \{[\alpha(I) \times \beta(I) \times IUUB\,(I)]\big|_{Si}\} \times \{A\,(I)\big|_{Si}\} \qquad \text{................ Equation (4)}$$

This brings the question to that of costs. As can be seen, the correct assessment of benefit from the information processing at the core *IS* model under consideration can be done *only* when, from the gross benefit as accruing under Equation (4), the costs of information processing are accounted for. What are these cost components then? Consistent with the individual information originating and processing nature of *IS*, it is suggested that these cost components are those of originating information "I" [denoted by $COST_{OI}$ (I)], of analyzing integrity quotient of A (I) [denoted by $COST_{ANAL}$ {A (I)}], and the opportunity cost of analyzing A (I) [denoted by $COST_{OPPORT}$ {A (I)}] [14]. Accordingly then the "net benefit" in the form of improvement in information use as accruing at the information processing stage (S_i) is as given in Equation (5).

$$\Delta IU(I)\big|_{Si} = [\{[\,\alpha\,(I) \times \beta(I) \times IUUB(I)]\big|s_i\} \times \{A(I)\big|s_i\}] - [COST_{OI}\,(I)\big|_{Si} + COST_{ANALY}\,\{A(I)\}\big|_{Si} + COST_{OPPORT}\,\{A(I)\}\big|_{Si}\,] \qquad \text{..... Equation (5)}$$

Accounting for dynamic situations characterizing the information flow, if one considerably simplifies the query at hand and assumes $\alpha(I)$ and $\beta(I)$ to be given (something not to be the case in real world problem solving), the functions IUUB(I) and A(I) having their own respective first order transients with corresponding steady state (ss) values (here of upper bound value for IUUB(I) and value equal to numerical one for A(I)), and assumes all cost functions to be exponentially increasing with time, then what emerges from Equation (5) is that the variable $\Delta IU(I)$ at the stage (S_i) under consideration *will* have a maximum value at a given time, and , among other things, *for* a given (what can be seen as an optimum, i.e., desired or, say, intended) value of integrity quotient "A". In other words there *is* an optimum I*I at which net increase in information use benefit is maximum (see Figure (1); Figure not to the scale); achieving that I*I (implying accuracy, consistency, and reliability - if they can be quantified) is a costly process; and, to meet the demands of competitive advantage, resource commitment for achieving improved I*I, preferably optimum I*I, is critical.

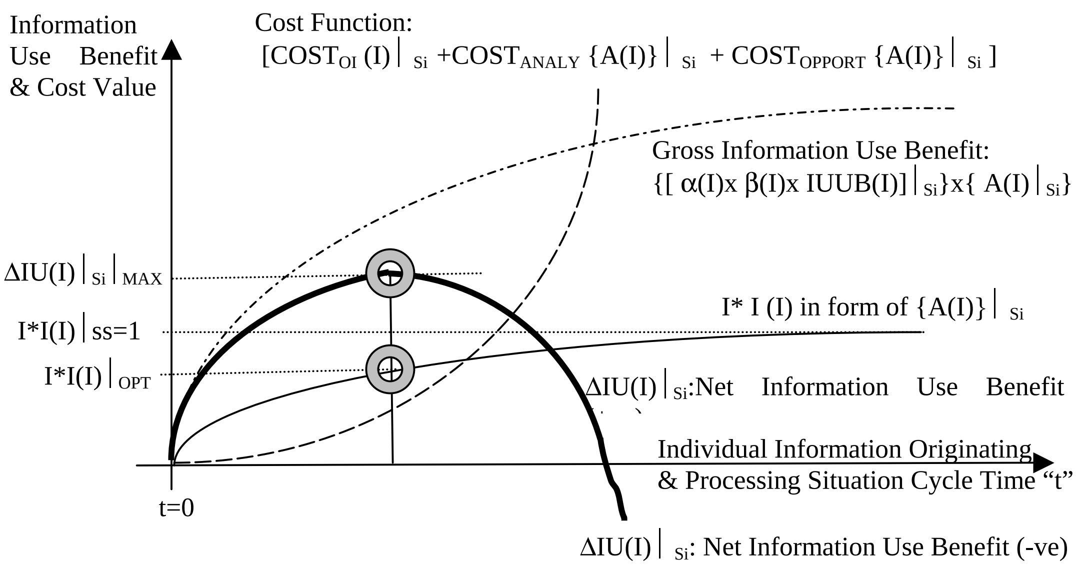

Figure (1): Cost-Benefit Analysis of Information Integrity

Equation (5) (with A(I) substituted by I*I(I)) thus gives the *IS* model of the continuous individual information originating and processing situation under uncertainty. Specifically by modeling information origination as a costly process, the model presents I*I (I) as a bottleneck and, therefore, as a resource for improved information use establishing analytically requirement for I*I; the desired or optimum I*I denoted as $I*I\ (I)|_{OPT}$ being that value of I*I, which maximizes ΔIU(I).

9. EQUATION FOR VALUE OF INFORMATION

For deriving equation for the value of information, one could begin by formulating generalized cost-benefit equation of I*I by putting Equation (2) in Equation (5); and the same is given by Equation (6).

$$\Delta IU(I) = [\{\alpha(I) \times \beta(I) \times IUUB(I)\} \times \{A(I) \times C(I) \times R1(I) \times R2(I)\}] - [COST_{OI}(I) + \{COST_{ANAL}(A(I)) + COST_{OPPOR}(A(I))\} + \{COST_{ANAL}(C(I)) + COST_{OPPOR}(C(I))\} + \{COST_{ANAL}(R1(I)) + COST_{OPPOR}(R1(I))\} + \{COST_{ANAL}(R2(I)) + COST_{OPPOR}(R2(I))\}] \quad \text{…Equation (6)}$$

We have not considered here quantification of I*I attributes. But for the purpose of investigation at hand, let us assume that $[0<A(I)\leq 1, 0<C(I)\leq 1, 0<R1(I)\leq 1, 0<R2(I)\leq 1]$. Then, it follows that $[0<[I*I\ (I)=\{A(I) \times C(I) \times R1(I) \times R2(I)\}]\leq 1]$ [7,8]. Information Integrity risk is then as given in Equation (7).

Information Integrity Risk = I*I Risk = {1 – [A(I) x C(I) x R1(I) x R2(I)] } …. . Equation (7)

The question now is what maximum possible information value the *IS* modeled by Equation (6) can make use of so as to improve trustworthiness of information "I"? Of course, if I*I (I) has value "1", then one can see all the determinants of I*I, that is, A(I), C(I), and R1(I) and R2(I) would be having value "1" each. In other words, there is a total confidence or trustworthiness in respect of information "I", and, hence, it can be said there is no use getting any further information, the additional information being of zero value. But what if I*I (I) <1? Then, the *IS* could, according to integrity estimate provided, process additional information "I_{ADDI}" in order to improve information use; the maximum improvement in information use being possible when I*I (I) is increased to value "1". Hence, the maximum possible value of I_{ADDI}, denoted by $I_{ADDIMAX}$, that the *IS* can usefully process is as given by Equation (8).

$$I_{ADDIMAX} = [\alpha\ (I) \times \beta\ (I) \times IUUB(I)] - \{[\alpha\ (I) \times \beta\ (I) \times IUUB(I)] \times [A(I) \times C(I) \times R1(I) \times R2(I)]\}$$
$$= [\alpha\ (I) \times \beta\ (I) \times IUUB(I)] \times \{1 - [A(I) \times C(I) \times R1(I) \times R2(I)]\} \quad \ldots\ldots\ \text{Equation (8)}$$

In other words, for an *IS*, maximally valuable information (when not accounting for cost components) would be that information which removes all risk of processing information less optimally, bringing the estimate of I*I (I) to value "1". However, the value of the additional information "I_{ADDI}", though useful, that the *IS* actually will process may not be the maximum possible. This is because of two reasons:(i) firstly, because I_{ADDI} may not have the requisite level of Information Integrity so as to increase point scales of A(I), C(I), and R1(I) and R2(I) to value of "1" each, which is necessary to bring I*I (I) to "1", and (ii) secondly, because there is a need to take into account the cost components in respect of originating I_{ADDI}, and of analysing its integrity and its opportunity cost for the purpose. This then leads in the manner of cost benefit analysis to the statement of value of information "I_{ADDI}" and the same is given in Equation (9).

$$I_{ADDI} = [I_{ADDIMAX} \times \{I{*}I \text{ of } I_{ADDI})\}] - [\text{Cost components concerning additional information } (I_{ADDI}) \text{ that can be usefully processed by } IS] \quad \ldots\ldots\text{Equation (9)}$$

Putting Equation (8) into Equation (9) and expanding for cost components, one gets Equation (10) giving value of additional information, I_{ADDI}, in terms of Information Integrity component of information "I".

$$I_{ADDI} = [\alpha\ (I) \times \beta\ (I) \times IUUB(I) \times (1- [A(I) \times C(I) \times R1(I) \times R2(I)]) \times (A(I_{ADDI}) \times C(I_{ADDI}) \times R1(I_{ADDI}) \times R2(I_{ADDI}))] - [COST_{OI}(I_{ADDI}) + \{COST_{ANAL}(A(I_{ADDI})) + COST_{OPPOR}(A(I_{ADDI}))\} + \{COST_{ANAL}(C(I_{ADDI})) + COST_{OPPOR}(C(I_{ADDI}))\} + \{COST_{ANAL}(R1(I_{ADDI})) + COST_{OPPOR}(R1(I_{ADDI}))\} + \{COST_{ANAL}(R2(I_{ADDI})) + COST_{OPPOR}(R2(I_{ADDI}))\}] \quad ..\text{Equation (10)}$$

Equation (10) requires calculation of I*I of I_{ADDI} — a further research investigation query, not pursued here. But, I*I of I_{ADDI} has its own attributes of accuracy, consistency, and reliability, which would need to be obtained. Secondly, Equation (10) is such that, apart from accounting for cost components in the manner of Equation (6), it expresses additional information value as a function of what is known about information "I" and what is actually obtained from additional information as it is processed to improve on integrity of information "I". And, finally, it is evident that, by controlling integrity of additional information, it is possible for *IS* to process higher value of useful additional information, once again reiterating the criticality of I*I (first time in respect of information "I" as in Equation (6) and, now, in respect of additional information "I_{ADDI}") for maximization of information *use*.

10. IMPROVEMENT IN VALUE OF INFORMATION INTEGRITY DUE TO ADDITIONAL INFORMATION

With value of additional information determined, the question to be answered is the relationship between I*I(I+I_{ADDI}) and I*I(I). We have I*I(I) ={A(I) x C(I) x R1(I) x R2(I)}(Equation (2)); where I*I risk is given by [1-{A(I) x C(I) x R1(I) x R2(I)}] (Equation (7)). Thus, as one considers I*I(I+I_{ADDI}), which is given by {A(I+I_{ADDI}) x C(I+I_{ADDI}) xR1(I+I_{ADDI}) xR2(I+I_{ADDI})}, one is considering improvements in integrity attribute values A(I), C(I), R1(I), R2(I) of information "I" due to values of integrity attributes of I_{ADDI}; and it is reasonable to suggest that each of integrity attributes of I_{ADDI} would have their respective determinants (in the manner of integrity of information on integrity). How does one proceed in that case? A workable approach is to consider a case when Consistency (C) and Reliability (R1) and (R2) attribute values for additional information are "1" each; that is there is complete confidence or trustworthiness in respect of Consistency and Reliability of I_{ADDI}. In such a situation, improvement in A(I) due to additional information "I_{ADDI}" is function of value of Accuracy attribute of I_{ADDI} denoted by $A(I_{ADDI})$. Given that, for [I*I(I) ={A(I) x C(I) x R1(I) x R2(I)}], I*I risk is [1- {A(I) x C(I) x R1(I) x R2(I)}]; it then emerges that possible improvement in I*I (I) is $\{A(I_{ADDI})$ x [1 – { A(I) x C(I) x R1(I) x R2(I)}]}. This gives $\{A(I+I_{ADDI}) \times C(I+I_{ADDI}) \times R1(I+I_{ADDI}) \times R2(I+I_{ADDI})\}$ as in Equation (11).

$$[A(I+I_{ADDI}) \times C(I+I_{ADDI}) \times R1(I+I_{ADDI}) \times R2(I+I_{ADDI})] = [A(I) \times C(I) \times R1(I) \times R2(I)] + \{A(I_{ADDI})\big|_{C(IADDI)=1,\ R1(IADDI)=1,\ R2\ (IADDI)=1} \times [1- \{A(I) \times C(I) \times R1(I) \times R2(I)\}]\} \quad \text{…Equation (11)}$$

Rearranging R.H.S and L.H.S terms in Equation (11) gives expression for $A(I_{ADDI})$ as in Equation (12).

$$\{A\ (I_{ADDI})\big|_{C\ (IADDI)=1,\ R1\ (IADDI)=1,\ R2\ (IADDI)=1}\} = \frac{[A(I+I_{ADDI}) \times C(I+I_{ADDI}) \times R1(I+I_{ADDI}) \times R2(I+I_{ADDI})] - [A(I) \times C(I) \times R1(I) \times R2(I)]}{[1- \{A(I) \times C(I) \times R1(I) \times R2(I)\}]} \quad \text{… Equation (12)}$$

But,

$$A(I+I_{ADDI})=A\ (I)+\Delta A\ (I),\ C(I+I_{ADDI})=C(I)+\Delta C(I),\ R1(I+I_{ADDI})=R1(I)+\Delta R1(I),\ R2(I+I_{ADDI})=R2\ (I)+\Delta R2\ (I) \quad \text{…Equation (13)}$$

Putting equations in Equation (13) in Equation (12), one gets equation (14).

$$\{A(I_{ADDI})\big|_{C\ (IADDI)=1,\ R1\ (IADDI)=1,\ R2\ (IADDI)=1}\} = \frac{[\{A\ (I) + \Delta\ A\ (I)\} \times \{C\ (I) + \Delta\ C\ (I)\} \times \{R1\ (I) + \Delta\ R1\ (I)\} \times \{R2\ (I) + \Delta\ R2\ (I)\}] - [A\ (I) \times C\ (I) \times R1\ (I) \times R2\ (I)]}{[1- \{A\ (I) \times C\ (I) \times R1\ (I) \times R2\ (I)\}]} \quad \text{…Equation (14)}$$

With A (I_{ADDI}) defined assuming $C(I_{ADDI})$, $R1(I_{ADDI})$ and $R2(I_{ADDI})$ each equal to "1", I*I attributes of information "I" would undergo improvements *only* along the Accuracy axis. This is an interesting visualization in that it encourages viewing I*I space with four dimensions in terms of: Accuracy (A) axis, Consistency (C) axis, and Reliability (R1) axis and Reliability (R2) axis. We are considering here {A (I)<1,C (I)<1,R1 (I)<1,R2 (I)<1}. Then, the information "I" has Information Integrity Risks (I*I Risk) in respect of Accuracy, Consistency, and Reliability attributes as shown in Equation (15).

(a)Accuracy (A) risk for "I" = {1- A(I)}, (b)Consistency (C) risk for "I"={1-C(I)},
(c)Reliability (R1) risk for "I"={1-R1(I)}, (d)Reliability (R2) risk for "I" ={1-R2(I)} ..Equation (15)

Due to additional information, accuracy attribute value for which is as shown in Equation (14), it is these Accuracy, Consistency, and Reliability risks of information "I" given in Equation (15), which are then to undergo reduction. And, as a result of the (extreme) case under consideration of $C(I_{ADDI})$, $R1(I_{ADDI})$, $R2(I_{ADDI})$ equal to "1", all these reductions in I*I attribute risk values for information "I", are to be effected by *only* accuracy components of I_{ADDI} . Accordingly, then, with respect to Equation (13), while maximum possible increments ΔA(I), ΔC(I), and ΔR1(I) and ΔR2(I) should facilitate removing all risks in A(I), C(I), R1(I), R2(I) raising their value to "1" each, in the case under consideration it is not so; the increments working out to be as shown in Equation (16).

$\Delta A(I) = AA(I_{ADDI}) \times \{1 - A(I)\}$, $AA(I_{ADDI}) \triangleq$ Accuracy component of $A(I_{ADDI})$ effecting A(I);
$\Delta C(I) = AC(I_{ADDI}) \times \{1 - C(I)\}$, $AC(I_{ADDI}) \triangleq$ Accuracy component of $A(I_{ADDI})$ effecting C(I);
$\Delta R1(I) = AR1(I_{ADDI}) \times \{1 - R1(I)\}$, $AR1(I_{ADDI}) \triangleq$ Accuracy component of $A(I_{ADDI})$ effecting R1(I);
$\Delta R2(I) = AR2(I_{ADDI}) \times \{1 - R2(I)\}$, $AR2(I_{ADDI}) \triangleq$ Accuracy component due to $A(I_{ADDI})$ effecting R2(I)
...Equation (16)

Putting equations in Equation (16) in Equation (14), one then gets Equation (17).

$$\{ A(I_{ADDI}) |_{C(IADDI)=1,\ R1(IADDI)=1,\ R2(IADDI)=1} \} = \frac{[\{A(I)+AA(I_{ADDI}) \times \{1-A(I)\}\} \times \{ C(I) + AC(I_{ADDI}) \times \{1-C(I)\}\} \times \{ R1(I) + AR1(I_{ADDI}) \times \{1-R1(I)\}\} \times \{ R2(I) + AR2(I_{ADDI}) \times \{1-R2(I)\}\}] - [A(I) \times C(I) \times R1(I) \times R2(I)]}{[1-\{A(I) \times C(I) \times R1(I) \times R2(I)\}]}$$

$$= \frac{\text{Net change in I*I due to Additional Information}}{\text{Maximum possible change in I*I due to additional information}} \quad \text{.....Equation (17)}$$

And rearranging the L. H. S. and R.H.S. terms, one gets Equation (18).

$$\text{Net improvement in I*I (I) due to } I_{ADDI} = \{ A(I_{ADDI}) |_{C(IADDI)=1,\ R1(IADDI)=1,\ R2(IADDI)=1} \} \times [1-\{A(I) \times C(I) \times R1(I) \times R2(I)\}] \quad \text{..Equation (18)}$$

Equations (17) and (18) give improvement in value of I*I in the form of $A(I_{ADDI})$ and the net improvement in I*I(I), respectively, as functions of determinants of previously accumulated information, in this case denoted by "I", and the additional information. Needless to say, similarly one will need to develop the net I*I improvement equations due to $C(I_{ADDI})$ and $R1(I_{ADDI})$ and $R2(I_{ADDI})$. This completes the investigation at hand by giving in the form of Equations (5) and (6) cost-benefit analysis of Information Integrity amenable to comparing of a set of integrity mechanisms {I*I(I)} and for selection of optimum integrity mechanism ($I^*I(I)|_{OPT}$) by the way of maximization of net information use quantum $\Delta IU(I)$ for competitive advantage. Further, Equation (10) gives the equation for value of information, and Equations (17) and (18) give improvement in the value of Information Integrity.

Of course, Equation (10) has cost components for which there is a need to develop equations. Also, Equations (5) and (6), (10), (17) and (18), for calculation of respective values, are all functions of I*I attribute and I*I values previously accumulated. Thus there is the issue of developing appropriate metrics and methods for quantification of I*I attributes. In fact there can be different formulations for even Equation (2) [8]. Further, investigation here has been mainly concerned with attributes of I*I for information value, i.e., for the content of information. Information Integrity is a systems concept [7, 8]. This implies similarly one should develop descriptions of accuracy, consistency, reliability in respect of process integrity and system integrity and in respect of design, development, implementation, and maintenance integrity. And, for all these descriptors also there is a need to detail equations as developed in this paper. All these and many other related aspects constitute further I*I research issues.

11. CONCLUSION

If business managers are found wanting in allocating resources to orient or develop their internal and external performance control systems emphasizing Information Integrity, it is because in the manner of cost-benefit analysis the business information system models currently at their disposal are not amenable to analytically compare two Information Integrity mechanisms and to select for competitive advantage one that optimizes the information *use*.

REFERENCES

[1] COSO – Committee of Sponsoring Organizations of the Treadway Commission, *Internal Control – Integrated Framework*. AICPA, Jersey City, 1992.

[2] Courtney R.H. and Ware W.H., What do we mean by Integrity? *Computers & Security*, (13) 1994, pp. 206-208.

[3] Delone W. H. and McLean E. R., Information Systems Success: The Quest for the Dependent Variable. *Information Systems Research*, 3 (1), 1992.

[4] Demski J. S., *Information Analysis.* Addison Wesley, Reading, MA, 1980.

[5] Dörner D., *The Logic of Failure* **-** *Recognizing and Avoiding Error in Complex situation*. Perseus Books, USA, 1996.

[6] Hogarth R. M. and Reder M. W. (1986). *Introduction: Perspectives from Economics and Psychology*. In 'Rational Choice — The Contrast between Economics and Psychology' edited by Hogarth R. M. and Reder M. W., The Chicago University Press, Chicago, 1986.

[7] Mandke V. V. and Nayar M. K., Information Integrity - A Structure for its Definition. *Proceedings of the 1997 MIT Conference on Information Quality* edited by Diane M. Strong and Beverly K. Kahn, Cambridge, Massachusetts, USA, October 25-26, 1997.

[8] Mandke V.V. and Nayar M.K., Implementing Information Integrity Technology- A Feedback Control System Approach. *IFIP TC11 WG11.5 Third Working Conference on IICIS, Amsterdam, The Netherlands, November 18-19, 1999*,Edited by Magaret E. van Biene-Hershey and Leon A.M. Strous, Kluwer Academic Publishers, USA.

[9] Mandke V. V., Nayar M. K. and Malik Kamna, Information Envelope and its Information Integrity Implications: For a complex, changing environment, modeling a generic business process as an integral to a closed loop information and control system characterized by uncertainty. *Proceedings of the 2001 MIT Conference on Information Quality* edited by Elizabeth M. Pierce and Raissa Katz-Haas, Cambridge, Massachusetts, USA, November 3-4, 2001.

[10] Mathews D. Q., *Design of the Management Information System.* Auerback Publishers, NY, 1971.

[11] Morton J. E. and Felix W. L. Jr., A Critique of Statement on Auditing Standards No. 55. *Accounting Horizons*, March 1991.

[12] O'shea G. F. G., Operating System Integrity. *Computers &Security*, 10 (1991), 443-465.

[13] Storey N., *Safety-Critical Computer Systems.* Addison-Wesley Longman, England, 1996.

[14] Tallberg A., *An Economic Framework for Information Integrity*. Forskningsrapporter Research Reports, Swedish School of Economics and Business Administration, Finland, 1999.

0-0-0-0-0