THE UNIVERSITY OF ARIZONA ®
GRADUATE COLLEGE

As members of the Final Examination Committee, we certify that we have read

the dissertation prepared by IRIT ASKIRA GELMAN

entitled Studies on the Informativeness, Value, and Cost, of Information and Information Systems

and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy

David E. Pingry — Date

Daniel Zeng — Date

Sudha Ram — Date

Date

Date

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copy of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

Dissertation Director David E. Pingry — Date

STATEMENT BY AUTHOR

TABLE OF CONTENTS

TABLE OF CONTENTS - *CONTINUED*

LIST OF ILLUSTRATIONS

LIST OF TABLES

ABSTRACT

The widely used database technology and more recent developments in networking and Web technologies are encouraging diversity in the utilization of existing data. Data are now routinely pooled from multiple systems and physical locations, and integrated in creative ways for various decision-making purposes. From a managerial perspective, however, there are growing concerns in regard to the quality of the output information, and the economic justification of costly investments in such technologies.

The major part of this dissertation addresses these concerns through formal studies on the quality and value of information, based on information economics (IE) theory.

The quality and value of information integration is studied from a standpoint that recognizes the fundamental role of information integration in information systems. The objective of this study is to create a domain-independent theoretical framework that can facilitate decision making on information integration. The framework classifies information integration situations using two information quality characteristics—*informativeness* and *dependence*—and links different conditions in terms of these characteristics with different predictions on the value of integration.

A second, related study centers on the questions of whether improving the accuracy of the input of an information system guarantees higher accuracy and economic value of its output, especially higher accuracy and economic value of forecasts. The study

offers sufficient conditions under which the answer to these questions is positive, and also presents counter examples that suggest conditions under which the answer is negative. The results point to a contextual factor that can affect accuracy both ways—positive or negative—which has been ignored by data quality theory. This factor is *dependence between errors*.

A third study considers a question related to software economics. Software economics theory equates software with code and directs that the supply of information be based entirely on demand patterns. However, an increasingly common custom in the software market to bundle the code with services indicates that a different model of cost and price may apply in many cases. Such model combines the information goods cost model with a service cost model. The study focuses on the question of the validity of such alternative model.

1. INTRODUCTION

The progress in communication technology in recent years has been accompanied by a spreading belief in the importance and value of information integration. Substantial financial investments in information integration technologies have been made, and today, data are routinely pooled from multiple systems and physical locations, and integrated in creative ways for various decision-making purposes (Stallings, 2003; Garcia-Molina et al., 2001).

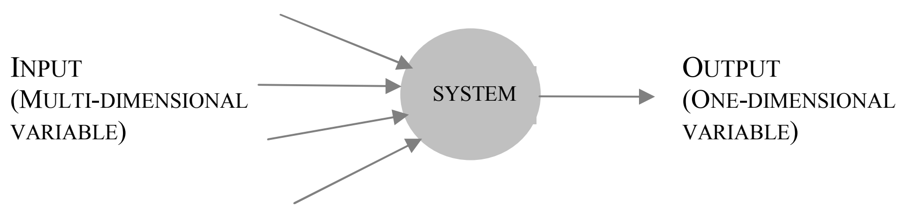
Information integration is usually explored in MIS literature from a technological perspective. For example, a popular database textbook (Garcia-Molina et al., 2001) defines information integration as “applications [that] take data that is stored in two or more databases (*information sources*) and build from them one large database, possibly virtual, containing information from all the sources, so the data can be queried as a unit.”

The technical challenge and high cost that characterize information systems integration and distributed information systems have motivated research studying the economic value of information integration (Goodhue et al., 1992). This is the main focus of this dissertation.

The dissertation consists of three studies. The objective of the first study is to create a domain-independent theoretical framework that can facilitate decision making on information integration. Such framework should enable (a) identification of critical factors that impact the value of integration of information sources and (b) prediction of the extent to which value can be extracted from integration of given information sources.

In agreement with the broad applicability that this investigation seeks, its interpretation of the notion of information integration emphasizes the typical *purpose* of an information

FIGURE 1.1: Information integration



integration technology, system, or product, rather than any specific choice of the former. It is assumed that information integration produces information through synthesis of multiple input information sources. In other words, the term information integration is associated with a function that maps the values of a multi-dimensional variable to the values of a one-dimensional variable (Figure 1.1). From this standpoint information integration designates, in effect, a fundamental role of information systems. Subsequently, a formal inquiry approaches the question of the economic value of information integration from an information quality perspective.

A second, related, study in this dissertation targets the question of whether improving the accuracy of the input of an information system guarantees higher accuracy and economic value of its output, especially higher accuracy and economic value of forecasts.

The colorful expression—“garbage in garbage out”—reflects a commonly accepted belief among the MIS community—the accuracy of the output of an information system is tightly and positively linked to the accuracy of its input. Yet, does an increase in input accuracy indeed guarantee an increase in output accuracy? A positive answer—along the spirit of “garbage in garbage out”—may seem obvious when the aspect of use is neglected, i.e., what purposes the output (and therefore the input) serves. In fact, it may be used by multiple entities, for a variety of purposes (Ballou and Tayi, 1999). Suppose that, in a decision-making context, sales data are used for estimating future demand. Does a lower number of errors in the sales data guarantee that the forecast of future demand will be more accurate?

Such questions are of great interest. Improving input accuracy can be a costly endeavor—it may require re-organization of work processes, acquisition of new technologies (new hardware, new software), etc. Therefore there is a need to assess the accuracy of the output given whatever use is made of the input, and interpret quality distinctions in economic terms.

MIS research has typically focused on circumstances in which the output supports only single activities, of an operational, rather than decision-making, nature (Ballou and Tayi, 1999). This research, in contrast, analyzes the changes in accuracy and economic value acknowledging the diverse use of data in practice, especially in decision-making.

Finally, this dissertation also addresses a question in the area of software economics. Current software economics theory—the information goods economics model and software economics literature—equates software with code and directs that the supply of information be based entirely on demand patterns. However, an increasingly common practice in the software market to bundle the code with services indicates that a different model of cost and price may apply in many cases. Such model combines the information goods cost model with a service cost model. The current investigation focuses on the question of the validity of such alternative model.

Each of the research questions, as well as the respective research method, is presented in further detail in the following sections.

1.1 DOES HIGHER DATA ACCURACY PRODUCE HIGHER FORECAST ACCURACY?

Is the relationship between the accuracy of an information system's input and the accuracy of its output "monotonic"? In other words—does higher accuracy of the input of an information system guarantee higher accuracy of its output? This question is the substance of an investigation covered in Chapter 3 and part of Chapter 4. A joined question is explored simultaneously, i.e., whether higher accuracy of the input of an information system guarantees higher economic value of its output.

1.1.1 Motivating managerial scenarios

Scenario #1

An industrial equipment producer developed a component that improves the efficiency of a piece of equipment that many of its clients use. Based on a survey of a sample of existing customers, it was estimated that p of the equipment units (e.g., p=0.5) in use by each client will be enhanced with the new component in the first year following its introduction to the market.

In order to get a basis for a more precise assessment of the demand among customers, marketing analysts turned to customer records. These included data about the number of equipment units currently in use by each customer. However, data in customer records were not completely accurate. The errors corresponded to one of two types: A few of the errors related to records that were out-of-date; a second subset corresponded to incomplete records. The number of units in use by the customer had been set in the latter case to a conservative default.

Management is aware of the poor quality of customer data, and may want to invest in better data quality. However, it may not be aware of certain additional factors linked to data quality. First, the records that are out-of-date are common among dynamic, fast-moving customer organizations. In these organizations more than p of the units will have the new component by the end of the first year. Second, incomplete customer records are common among customers that suffer from weakness of control and management (e.g., they find it difficult to provide the specified data). The same weakness is affecting their decision processes, which are relatively slow. Consequently, significantly less than p of the units will have the new component in those organizations within the first year after its introduction.

Suppose that the forecasts desired by the producer are obtained by multiplying the number of units in use by p. Will more accurate customer records imply more accurate forecasts?

This scenario is consistent with a condition suggested by the research, under which the answer can be negative: *the input error and the forecast error are correlated*. Hence, depending on the magnitude of such correlation, the conservative (i.e., low) default values applied to the incomplete records can generate more accurate forecasts of future demand among the "slow" customers than the actual numbers. Similarly, present records for the "dynamic" customers can result in more accurate forecasts if the errors in such records portray systematically higher recorded values than the actual numbers (for example, if the actual number of units in use has dropped systematically in a shrinking market).

1.1.2 Current state of research

The influence of higher accuracy of an information system's input on the accuracy and economic value of its output is addressed by a number of studies in MIS. Related questions about measurement-errors have also been studied in econometrics and other social sciences.

Empirical MIS studies often make specific assumptions on the nature of the integration algorithm, e.g., a linear regression model (Klein and Rossin, 1999) or a neural network algorithm (Hwarng 2001). Their findings have been inconclusive. Data quality research has also produced a more general theory. This theory suggests that accuracy is an intrinsic properties of data, i.e., it does not vary across different applications that use the data (e.g., Wand and Wang, 1996). It also suggests that the accuracy of a calculated value is a deterministic, monotonically increasing function of the accuracy measures of the respective inputs, i.e., higher input accuracy implies higher accuracy of the information system output (Ballou et al., 1998).

This study re-examines such theory in situations where the relationships of the data to phenomena of interest are either stochastic or deterministic, rather than the purely deterministic relationship assumed by earlier work.

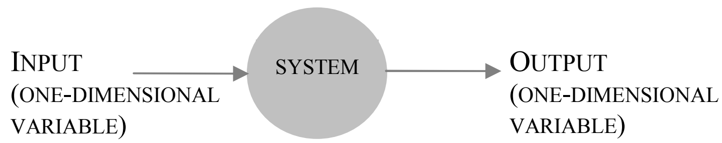
1.1.3 Research approach

Results are achieved through formal analysis. For the most part, the analysis applies the information structure (IS) model and respective theory (e.g., Blackwell, 1953; Marschak, 1971; McGuire, 1986). The IS model, rooted in statistical decision theory and game theory, has been used in many studies in MIS and other research areas.

Accuracy and a more general notion that refers both to the accuracy and resolution of the information, i.e., informativeness, are defined formally, using Blackwell's sufficiency criterion (Blackwell, 1951; Blackwell, 1953). Subsequently, Blackwell's Theorem (Blackwell, 1953; Marschak, 1971) establishes that higher accuracy (/informativeness) is equivalent to superior economic value. This equivalence implies that research results are equally valid for accuracy (/informativeness) and economic value.

Chapter 3 investigates the research questions under the assumption that the information system uses a single input source for generating the output information. In other words, the information system is viewed as a function that maps the values of a one-dimensional variable to the values of a one-dimensional variable (Figure 1.2). Chapter 4 considers these questions under the assumption that the information system produces information

FIGURE 1.2: Output information is generated based on a single input data source



through synthesis of multiple input sources (Figure 1.1). The analysis identifies conditions on the system's input and output that imply "monotonicity", or, alternatively, "non-monotonicity".

Part of the analysis in Chapter 3 applies further statistical theory. It interrogates a type of *non-classical measurement errors* in the context of a linear regression model, for directing to a class of situations in which the answer to the research questions is negative. Social sciences literature normally handles data quality issues based on a set of assumptions that are known as "classical measurement errors". However, a growing body of research on non-classical measurement errors is motivated by the understanding that the classical assumptions often "reflect convenience rather then conviction" (Bound et al., 1994).

1.2 ASSESSING THE INFORMATIVENESS AND VALUE OF INFORMATION INTEGRATION

The inquiry reported in Chapter 4 concentrates on the accuracy, informativeness, and economic value of information integration under various conditions.

The framework in use classifies information integration situations based on two information quality characteristics—*informativeness* and *dependence*. More precisely, the study categorizes integration scenarios using two tests that take for granted the state set that is targeted by the information. (1) Whether or not the input information sources can be ranked in terms of their *informativeness* for the problem at hand, and (2) whether or not the information sources are statistically *state-conditionally independent*. Four categories are created in this way.

Research questions include: What is the connection between the outcome of the informativeness test and the informativeness of the output of the integration? What is the connection between the dependence test outcome and the informativeness of the output of the integration? What is the combined influence of the dependence and informativeness tests outcomes on the informativeness of the output of the integration?

Similar questions are examined with respect to the accuracy, and the economic value, of information integration.

The research on information integration also complements the investigation on whether the relationship between the accuracy of the information system's input and the accuracy of its output is monotonic. It scans the questions under the assumption that the information system produces information through synthesis of multiple input sources.

1.2.1 Motivating managerial scenarios

Scenario #2

Suppose that, due to historic or other reasons, enterprise data, e.g., enterprise sales data, are recorded independently by two or more units in the organization. The firm is not satisfied with the quality of the data. As a result, it considers developing a procedure that will use existing data sets to synthesize an output set. Management hopes that the output will have strictly higher informativeness than the inputs. To what extent is management's hope substantiated?

The results of this study show that in many cases integration can generate strictly higher informativeness than those of the input sources, but sometimes it cannot. Suppose, for example, that errors form a completely random noise. In that case integration always has the potential to produce output with strictly higher informativeness. But the informativeness of the output would nonetheless be strongly bounded, such that it would be impossible to reach perfect information quality.

Notably, this case, and more generally, whenever sources are state-conditionally independent, point to a source selection strategy that may assist if there is a limit on the number of sources that can take part in the integration: integrate the most informative sources. This strategy has the potential to yield the highest informativeness and economic value under the assumed condition.

Scenario #3

In a chain of retail stores, predictions of sales in each store used to be made separately, i.e., a store's sales data served as a basis for the estimation of its future sales. Lately, however, management decided to try to use the data from a larger sample of the stores in forecasts of sales in each store. The underlying belief is that the combined predictive accuracy will be higher than that of the data of one store. Should the organization follow the strategy of selecting those stores whose data provide the best predictions for the given store?

Not necessarily. This strategy is recommended in scenario #2 based on a condition that may not be met under this scenario, i.e., state-conditional independence. Another strategy that may work better in both scenarios if that independence condition is not satisfied, or if the most informative sources cannot be identified, is to search for "complementary expertise" among information sources. One type of such complementarity may be discovered through a search for highly informative data "slices". Specifically, even when an information source is not highly informative as a whole, subsets of its overall value range may still be highly informative. When different information sources have complementary "slices" of this kind, their integration can reach perfect information quality and maximal value even if, individually, each has poor informativeness overall.

Scenario #4

Suppose that, under either scenario #2 or scenario #3, an analyst expresses the belief that "independence is the most valuable property". He or she suggests trying to identify sources that are state-conditionally independent, and integrating the most informative among them. Is this integration strategy always superior when it is feasible?

In terms of optimality of output informativeness and economic value, independence is not always “the most valuable property”. In fact, strong conditional dependence can form a second type of “complementary expertise” among information sources, which can be very beneficial for informativeness, and can even produce perfect information and maximal economic value.

1.2.2 Current state of research

The economics of information integration has attracted research in MIS and related fields, based on transaction cost theory, agency theory, organizational information processing theory, and other theories (e.g., Malone et al., 1987; Bakos, 1991; Gurbaxani and Whang, 1991; Clemons and Row, 1992; Goodhue et al., 1992).

Unlike those streams of research which make strong assumptions on the decision-making domain of information integration, e.g., supply-chain scenarios, the present inquiry aims to develop a theory that can help decision making on information integration regardless of domain. The inquiry centers on information quality as a factor that determines the economic value of information integration. It is closely related to analytical research that accounts for information quality in detail, and makes minimal assumptions on a decision-making or problem-solving environment. Such research has been conducted in diverse areas like political science, statistics, econometric forecasting, and computer science. The history of that literature goes back to the work of the French mathematician Marquis de Condorcet in the eighteenth century.

The contribution of this investigation in view of similar studies is related to the unique operationalization of the problem, in particular, the utilization of the information structure (IS) model and associated theory.

1.2.3 Research approach

Information integration is defined formally. The analysis refers to the notion of "more informative", and a stronger variation on this notion, "strictly more informative", introduced in Chapter 4.

Guided by relevant literature, the framework that is adopted here for describing integration scenarios is based on two information quality characteristics—informativeness and dependence. Given any of the four categories created accordingly, the following criteria on the informativeness of the outcome of information integration are examined. (1) Whether or not the informativeness of the outcome of integration is strictly higher than those of any of the inputs, (2) whether or not integration can produce perfect information, and (3) whether or not higher input informativeness implies higher informativeness of the outcome of integration (monotonicity).

1.3 SOFTWARE ECONOMICS: DOES IT FOLLOW THE INFORMATION GOODS MODEL?

An inquiry on software economics in Chapter 5 looks into the question whether the price of software is affected by per-unit cost caused by a service component.

1.3.1 Current state of research

A recent commentary by Cusumano (2003) points to the economic importance of services for software product companies. In general, services increasingly form a major share of the software industry. However, despite this situation, the economic implications of the demand and supply of services for the industry are not well understood at this stage.

Code and service are, in fact, highly complementary. In many cases software is actually sold as a package that consists of code and one or more “free” or fee-based services, sometimes including hardware (ASP, 2002; Softletter & ASP, 2000; Dover, 2002). A conception of software as a bundle that consists of code and services, and possibly also hardware, may enrich the theory that addresses the supplier’s side, which currently interprets software mainly as code and ignores the market reality.

I examine this bundling view of the economics of software in a specific domain—the basic propositions of the popular information goods model. The information goods model suggests that software has high, sunken, first-copy cost, and small, nearly zero, per-copy cost of production. Therefore, pricing based on the marginal cost of production is bound to end in losses; pricing must be based on the buyer’s value only. In contrast, the cost of a package that combines code with service is affected by the properties of the cost of service. This cost has a large variable component, caused by labor requirements.

1.3.2 Research approach

Data was collected from a variety of sources in the industry. These included general industry associations’ publications, consulting companies’ publications, industry periodicals, web sites of software vendors, financial reports of software vendors. Data collection also included direct interaction with people in the industry (industry experts, software vendor employees).

In order to produce highly valid results, the data collection process adhered to principles that are often applied for that purpose in research as well as in practice. First, source reputation was an important criterion in the choice of information sources. Information indicating to the popularity and reputation of information sources was taken into account. Second, data obtained from multiple sources was compared, and inconsistent data were checked more deeply and/or dropped (triangulation). Finally, two industry experts reviewed the output. This exchange was repeated at different stages of the study.

1.4 ORGANIZATION OF THE DISSERTATION

- The dissertation is structured in the following way. Chapter 2 reviews the literature that is relevant to each of the studies. Chapter 3, “Does higher data accuracy produce higher forecast accuracy?” explores the research questions as described in Section 1.1. Chapter 4, “Assessing the informativeness and value of information integration”, addresses the research questions as described in Section 1.2. Chapter 5, “Software economics: Does it follow the information goods model?” focuses on the questions as described in Section 1.3. Chapter 6 summarizes the results and contribution of this research, and recommends further research directions. An appendix provides the complete proof of Theorem 4.5.1.

-

2. LITERATURE REVIEW

2.1 THEORETICAL FOUNDATIONS

The foundations of the information structure (IS) model and related theory were developed in statistical decision theory (e.g., Savage, 1954) and game theory (Von Neumann and Morgenstern, 1944). Marschak was among the first to develop a systematic theory of the economic value of information (Marschak, 1971).

Computer-based information systems, expert opinion, market surveys, etc., are all viewed under the IS model as forms of additional information that a decision-maker may seek. Specifically, a decision-maker is assumed to have prior information that affects his or her payoff. This information takes the shape of a probability distribution of the state of the world. Such information implies a preferred strategy—a feasible strategy that promises optimal expected payoff. However, additional information can change the objective or perceived probabilities of the possible states. A different strategy for obtaining an optimal expected payoff (which may, or may not, be the same payoff as the maximal expected payoff without the additional information) might then be preferred. In other words, additional information has the power to affect the decision-maker's decisions and expected payoff. The a priori value of the additional information (before getting the actual information) is calculated as the maximal expected payoff over all the possible "signals", minus the maximal expected payoff without additional information.

Definitions of an *information structure*—the chosen model of information— as well as the IS model, *informativeness*, *sufficiency*, and *Blackwell's Theorem*, are given next. The term sufficiency, proposed by David Blackwell (1951; 1953), and the term informativeness, proposed by Bohnenblust, Shapley, and Sherman (1949) form the basis for the notions of accuracy and informativeness that serve in this investigation. Sherman (1951), Stein (1951), Blackwell (1953), Boll (1955), and others proved the equivalence of sufficiency and informativeness under various assumptions. Notably, such equivalence links a purely statistical concept (sufficiency) with an economic notion (informativeness).

The definitions in this section follow McGuire (1986), and Marschak (1971). They form a formal starting point for the dissertation.

Definition (2.1.1): Information structure (IS). Let S denote a finite set of *states of the world*, and Y a finite set of *signals*. The elements of each of these sets are mutually exclusive and exhaustive, that is, one and only one state will occur, and one and only one signal will be observed. An *information structure* $\boldsymbol{A}:S\text{x}Y\rightarrow[0,1]$ is described by a Markov matrix where the value in row *i*, column *j*, shows the probability that a signal $y_j \in Y$ will be produced given that a state $s_i \in S$ occurs.◙

A Markov matrix is a matrix whose entries are nonnegative and whose rows (or columns) each sum to one.

An information structure is termed by Blackwell "experiment". Marschak and Radner introduced the term "information structure".

Definition (2.1.2): Information structure (IS) Model. Let S denote a finite set of *states of the world*, Y a finite set of *signals*, and D the decision-maker's finite set of feasible *actions*. The elements of each of these sets are assumed to be mutually exclusive and exhaustive.

Let $\boldsymbol{p}$ be a vector of *a priori* probabilities of the states of the world, $\boldsymbol{A}$ an IS as defined above, $\boldsymbol{R}:Y\mathrm{x}D \rightarrow[0,1]$, describing the decision-maker's *decision rule*, is given by a Markov matrix in which the value in row i column j is the probability that an action $d_j \in D$ will be adopted, given that the observed signal is $y_i \in Y$.

Finally, $\boldsymbol{U}$ is the *payoff* function $\boldsymbol{U}:D\mathrm{x}S \rightarrow \Re$, associating a payoff to each pair of strategy and state of nature. $\boldsymbol{U}_{i,j}$ is the payoff if an action $d_i \in D$ is adopted when the actual state is $s_j \in S$.

The expected payoff given $\boldsymbol{p}$, $\boldsymbol{A}$, $\boldsymbol{R}$, $\boldsymbol{U}$, is $\mathrm{tr}(\boldsymbol{ARUp'})=\Sigma_{s,y,d}\boldsymbol{U}(d,s)\boldsymbol{p}(s)\boldsymbol{A}(s,y)\boldsymbol{R}(y,d)$. "tr" refers to the mathematical trace operator, and $\boldsymbol{p'}$ stands for a square matrix containing the elements of $\boldsymbol{p}$ in its main diagonal.◙

A decision maker is assumed to choose a decision rule that maximizes her expected payoff, i.e., $\mathrm{Max}_{R}\ \mathrm{tr}(\boldsymbol{ARUp'})$.

Definition (2.1.3): Sufficiency. Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be two ISs defined on $S\mathrm{x}Y_A$ and $S\mathrm{x}Y_B$, respectively. $\boldsymbol{A}$ is *sufficient* for $\boldsymbol{B}$ if there exists a Markov matrix $\boldsymbol{M}$ such that $\boldsymbol{AM}=\boldsymbol{B}$. ◙

The ordering relation that the sufficiency criterion establishes over the set of ISs defined on S is not total. (Not every pair of ISs can be ordered.)

Marschak and Miyasawa (1968) introduced the term "garbling" in relation to Blackwell's sufficiency. $\boldsymbol{M}$ is interpreted as a "garbling" matrix. The signal produced by the information source modeled by $\boldsymbol{A}$ is "garbled" so that the probability distribution of the outcome matches $\boldsymbol{B}$.

Definition (2.1.4): Informativeness. Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be two ISs defined on $S\mathrm{x}Y_A$ and $S\mathrm{x}Y_B$, respectively. Then, $\boldsymbol{A}$ is *more informative* or *generally more informative* than $\boldsymbol{B}$ if for any set of actions D, payoff matrix $\boldsymbol{U}$, and vector $\boldsymbol{p}$ of a priori probabilities, $\mathrm{Max}_{R_B}\mathrm{tr}(\boldsymbol{BR_B Up'}) \leq \mathrm{Max}_{R_A}\mathrm{tr}(\boldsymbol{AR_A Up'})$, where $\boldsymbol{R}_A$, $\boldsymbol{R}_B$, denote the decision-maker's decision rule defined on $Y_A\mathrm{x}D$, $Y_B\mathrm{x}D$, respectively.◙

"Generally more informative" is a transitive and reflexive relation (Marschak, 1971).

Blackwell's Theorem (2.1.1). Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be two ISs defined on $S\mathrm{x}Y_A$ and $S\mathrm{x}Y_B$, respectively. Then, $\boldsymbol{A}$ is generally more informative than $\boldsymbol{B}$ if and only if $\boldsymbol{A}$ is sufficient for $\boldsymbol{B}$. ◙

Blackwell's theorem offers a practical tool for ranking the economic value of ISs, through the equivalence that it establishes. The ranking that it enables is not total, but it has the advantage that it is valid regardless of major contextual factors, i.e., the decision-maker's payoff function, action set, and prior probability distribution of the state. The IS model does not take cost into account.

A few MIS inquiries have applied the IS model to analyze the value of information under distinct decision-making settings. This dissertation benefits mainly from the work of Ahituv and Ronen (1988). MIS researchers have often expressed the hope that information economics will establish a foundation for a core formal theory of MIS. For example, a review of information economics (Kleijnen, 1980) concludes: "In the field of information systems many intuitive ideas (rules of thumb) are around. Using the IE [information economics] model, one can prove some of these ideas to be true…Information economics is a methodology that until now has been applied mainly by economists and not by computer and information scientists… It remains to be seen whether students of IS [information systems] will be able to apply the IE framework and calculus to their systems." However, these hopes have not materialized so far (e.g., Robey, 1996). (See Kleijnen, 1980, or Repo, 1989, for a discussion on some difficulties with this theory.)

2.2 DOES HIGHER DATA ACCURACY PRODUCE HIGHER FORECAST ACCURACY?

Numerous MIS studies acknowledge the understanding that, ignoring costs, improvements in data accuracy are always beneficial. The following review aims at studies that investigate such ideas. In addition I provide a brief background on the notion of non-classical measurement-errors, since part of this research analyzes the effect of certain non-classical measurement-errors.

2.2.1 The relationship between the accuracy of an information system's input and the accuracy of its output

Wand and Wang (1996) propose ontological foundations for data quality dimensions, oriented towards system design and data production. They concentrate on intrinsic quality dimensions, i.e., dimensions that are use-independent or equivalent across applications. Their work provides a theoretical basis for the assumption that accuracy is an intrinsic property of data.

According to Wand and Wang's ontological proposition, both real world and information systems are systems, described in terms of states and laws. A good information system should be a "proper representation" of a real-world system. In a proper representation there is an exhaustive mapping from the lawful state space of the real-world system to that of the information system, which allows unique inference of the state of the real world from the state of the information system.

Wand and Wang suggest that intrinsic data quality dimensions, including accuracy, designate types of deviations of the information system from the conditions of a proper representation. In particular, inaccuracy corresponds to a situation in which a valid but incorrect state is inferred by the information system. The intrinsic nature is apparently derived from the assumptions of a single real-world system and the uniqueness of inference that defines a proper representation.

Ballou and Pazer (1985), and later, Ballou et al. (1998), offer general formulas for the assessment of accuracy and other quality attributes of an information system's calculated values. The error in the output is equal to the sum of the error in processing and the error due to input inaccuracy (Ballou and Pazer, 1985), or to some other monotonically increasing function of measures of these two error types (Ballou et al., 1998). The formulas imply that for any given processing error, the error in the output is a deterministic function of the input values and their deviations from the actual values (Ballou and Pazer, 1985), especially a monotonic increasing function of the measures of accuracy of the inputs (Ballou et al., 1998).

The model offered by Ballou and Pazer (1985) has been verified by a series of experiments (Ballou et al., 1987). The experiments assess the impact of errors in the data on forecasting model outputs in a spreadsheet environment. The benchmark values that the experiments employ for the assessment are forecast values generated by models that were constructed using error-free data. Hence, forecasting model outputs are not compared to the real world outcome, but rather, to the benchmark forecasts. The experiments show that errors in the data affect the output forecasts substantially, and they also affect the choice of the model type that provides the best fit to the data.

Mukhopadhyay and Cooper (1992) analyze the relationship between MIS output accuracy and decision accuracy in the context of an assessment of the usefulness of microeconomic production theory for MIS research. This relationship, which is comparable to the relationship between input accuracy and output accuracy discussed here, is analyzed through an inventory control decision-making problem. Their results confirm the microeconomic production theoretic view, according to which such relationship is positive, with diminishing marginal influence.

Empirical research that examined the relationship between input accuracy and output accuracy in non-deterministic settings shows diverse, mixed results. The results also raise question about the role of the chosen algorithm in such relationship, which is outside the current focus.

Bansal et al. (1993) compare the effect of errors in test data on the accuracy of linear regression and neural network models, forecasting the prepayment rate of mortgage-backed security portfolios. Error magnitude is shown to have a significant effect on the predictive accuracy of both the linear regression and the neural network models, and on a measure of payoff for the linear regression. Error rate has a significant effect on the predictive accuracy and payoff measure for the linear regression, but has no effect on the product of the neural network model.

Klein and Rossin (1999a) investigate the influence of errors on the prediction accuracy of linear regression models in forecasting the net asset value of mutual funds. One experiment set examines the influence of error rate and error magnitude in test data on the prediction accuracy of the model. The results demonstrate significant negative influence of both error rate and error magnitude. A second set of experiments examines the effect of error rate and error magnitude in data that serves for the construction of the regression model. These experiments show significant positive influence; both higher error rate and higher error magnitude increase the predictive accuracy compared to error-free data.

Klein and Rossin (1999b) conduct similar experiments with back-propagation neural network models. The first set of experiments shows that error rate and error magnitude in test data are negatively related to the predictive accuracy of the model. The second experiment set shows that a growing error magnitude in the training data has negative influence on the predictive accuracy of the resultant neural network. But a moderate error rate in the training data may have positive effect on the predictive accuracy compared to error-free data.

Hwarng (2001) studies the effect of noise in training data on the accuracy of a back-propagation neural network model in time-series forecasts. The results of the study vary. Some results show positive relationship, some negative relationship, and others do not indicate a relationship at all. The authors characterize the conditions in which data smoothing does/does not improve the prediction accuracy.

2.2.2 Non-classical measurement errors

Research on measurement-errors dates back to the 30's (Frisch, 1934). Studies in statistics, econometrics, and other social sciences typically employ the "classical measurement-errors" model. Take for example a simple linear regression model that describes the true relationship between two variables, x^* and y^*:

$$y^* = \beta x^* + \varepsilon$$

ε—the "stochastic disturbance"—is a random variable such that $\varepsilon \sim N(0,\sigma_\varepsilon^2)$, and is independent of x^*. Suppose, however, that x^* is unobserved, and an observed variable, x, is such that:

$$x = x^* + u$$

where u denotes the error in measuring x^*. Suppose also that y^* is unobserved, and an observed variable, y, is such that:

$$y = y^* + v$$

where v denotes the error in measuring y^*. Then, the classical measurement errors framework assumes the following (e.g., Fuller, 1991, Bound et al., 1994):

- Measurement errors have zero mean, $E(u)=0$, $E(v)=0$, i.e., the expected value of the mismeasured variables is equal to the expected value of the true measure.

- Measurement errors are independent of the variables of interest, i.e., each of u, v, is independent of x^*, y^*.

- The measurement errors and the stochastic disturbance (u, v, ε) are all mutually independent.

Researchers admit, however, that the classical assumptions often "reflect convenience rather then conviction" (Bound et al., 1994). These are strong assumptions, such that errors need not comply with them in general. The literature on measurement errors cites theoretical considerations that imply the existence of violations (e.g., errors in a binary valued variable are always negatively correlated with the true value, see Aigner, 1973). In addition, a growing number of empirical studies shows that in many real world instances the classical assumptions are violated. For example, when income data is provided by people, they misreport their income so that the errors are negatively correlated with job tenure, and positively correlated with education (Bound et al., 1994). People systematically overstate their education and the errors are correlated with the true measure, with other variables (e.g., race, gender), and with the residual in an earning regression (e.g., Black et al., 2003). Other non-classical measurement errors result from a weakness of memory of people. Researchers in this area aim to identify the behavior of such errors, reduce the error-rate, and/or account for the errors.

2.3 ASSESSING THE INFORMATIVENESS AND VALUE OF INFORMATION INTEGRATION

Information integration has been investigated under numerous terms, e.g., Condorcet's Jury theorem, expert resolution, feature selection, multisensor fusion, ensemble learning. Research has been conducted in areas as remote as statistics, computer science, political science, econometric forecasting, and more. Often, similar findings recur in different fields. Furthermore, measures of informativeness and dependence as well as other modeling assumptions vary greatly. For example, measures of dependence include the correlation coefficient (e.g., Frantzus 1967), "fraction of same errors" (Ali and Pazzani, 1996), Q statistic (Kuncheva et al. 2000), and others. Or, in regard to the integration mechanism, some studies base this mechanism on majority voting (e.g., Ladha 1992; Ladha 1995), others on Bayes rule (e.g., Clemen and Winkler, 1985), and so on.

Table 2.1 shows a list of relevant studies accompanied by the problem domains that motivated them. The list is organized by categories that will be clarified next.

TABLE 2.1: Classification of research on information integration

Unconditional independence	State-conditional independence
CONDORCET (1785) —GROUP ACCURACY	COCHRAN (1962) —LINEAR DISCRIMINANT FUNCTIONS
BARABASH (1965) —— FEATURE SELECTION IN PATTERN RECOGNITION	BARABASH (1965) —— FEATURE SELECTION IN PATTERN RECOGNITION
GROFMAN, OWEN, AND FELD (1983)—GROUP ACCURACY	ELASHOFF ET AL. (1967)—FEATURE SELECTION IN PATTERN RECOGNITION
NITZAN AND PAROUSH (1982); NITZAN AND PAROUSH (1984)—GROUP ACCURACY	FRANTSUZ (1967)—FEATURE SELECTION IN PATTERN RECOGNITION
YOUNG (1988) —GROUP ACCURACY	TOUSSAINT (1971)—FEATURE SELECTION IN PATTERN RECOGNITION
HANSEN AND SALAMON (1990) —NEURAL NETWORK ENSEMBLES	COVER (1974) —FEATURE SELECTION IN PATTERN RECOGNITION
LADHA (1992); LADHA (1995)—GROUP ACCURACY	COVER AND VAN CAMPENHOUT (1977) —FEATURE SELECTION IN PATTERN RECOGNITION
ALI AND PAZZANI (1996) —ENSEMBLE LEARNING	FANG (1979) — FEATURE SELECTION IN PATTERN RECOGNITION
KUNCHEVA ET AL. (2003)—ENSEMBLE LEARNING	CLEMEN AND WINKLER (1985) —ECONOMETRIC FORECASTING
	AHITUV AND RONEN (1988) —MANAGEMENT INFORMATION SYSTEMS
	JACOBS (1997) —NEURAL NETWORK ENSEMBLES
	LIU AND YAO (1998) —NEURAL NETWORK ENSEMBLES
	YOON AND SOMPOLINSKY (1999) —BIOLOGICAL NEURAL SYSTEMS
	ABBOTT AND DAYAN (1999) —BIOLOGICAL NEURAL SYSTEMS

Instead of the “ordinary” unconditional statistical independence and dependence, various studies assume *class-conditional independence* (and/or *class-conditional dependence*). The latter is also known as *state-conditional independence* (/*state-conditional* *dependence*). While unconditional independence requires that the probability that a random variable takes a particular value does not depend on the value of another variable, class-conditional independence requires that, *given any class*, the probability that a random variable takes a particular value does not depend on the value of another variable. When the probability distribution given different states is fixed these two types of statistical independence overlap. However, in the more general case they are usually contradictory (Barabash, 1965; Hussain, 1972).

2.3.1 Integration under unconditional dependence/independence assumptions

The classic *Jury theorem* (Condorcet, 1785), viewed as theoretical support for group decision-making and democracy, is an early theory of information integration. The theorem suggests that when group members are independent, each member judges correctly between a pair of alternatives with probability $p>0.5$, and the integration mechanism applies a majority rule, the probability that the output is correct, P_n, is higher than p ($P_n>p$). Group accuracy approaches perfect information as the number of members, n, grows to infinity ($P_n \rightarrow 1$ as $n \rightarrow \infty$).

Research on the integration of unconditionally dependent/independent sources shows, under varying additional conditions, that:

- When sources are independent the outcome of integration has higher accuracy than that of any individual input (Condorcet, 1785; Grofman et al., 1983; Nitzan and Paroush, 1982; Young 1988; Hansen and Salamon, 1990).

- When sources are independent accuracy approaches perfect information quality as the number of input sources grows to infinity (Condorcet, 1785; Grofman et al., 1983; Hansen and Salamon, 1990).

- As the accuracy values of the individual inputs grow higher, the accuracy of the outcome of integration grows higher too; the best inputs generate the best integration outcome (Grofman et al.,1983).

- Negative correlation enhances the accuracy of the output compared to independent information sources (Ladha 1992; Ladha 1995; Kuncheva et al. 2003).

- Positive correlation between sources has detrimental effect on accuracy compared to independent information sources (Nitzan and Paroush, 1984; Ladha 1992; Ladha 1995; Ali and Pazzani, 1996).

2.3.2 Integration under state-conditional dependence/independence assumptions

The work of Ahituv and Ronen (1988) in MIS establishes a fundamental result based on the IS model—the informativeness of the product of integration is not inferior to that of any of the input sources. Their work is portrayed in more detail in Chapter 4.

- Clemen and Winkler (1985) quantify the accuracy and value of state-conditionally independent versus state conditionally dependent sources. Their analysis shows that the integration of state-conditionally independent sources can always have superior accuracy compared to the sources, but it never reaches perfect information quality. In addition, the best inputs generate the best integration. Moderate positive correlation has dramatic negative effect compared to conditionally independent sources. Negative correlation and extreme positive correlation can have positive effect on the accuracy of the integration.

- The findings of Cochran, (1962), Yoon and Sompolinsky (1999), Jacobs (1997), and Liu and Yao (1998) are consistent, for the most part, with Clemen and Winkler (1985).

- Research on the integration of conditionally dependent / independent sources has partly centered on the problem of *feature selection*: given a set of *d* features, select a subset of size *m* that leads to the smallest classification error (Jain et al., 2000). Since an exhaustive analysis of all the different feature subsets is practically impossible even with moderate problem sizes due to computational limitations, research focuses on accuracy as well as search efficiency. The discoveries in this area suggest, for the most part, an uneasy message in regard to the quest for an efficient search.

Frantsuz (1967) analyzes an example that indicates that when features are strongly class-conditionally dependent, then, except for special cases, the joint informativeness of strongly conditionally dependent sources increases considerably compared to the sum of the informativness of the individual sources. Sources that have low informativeness in themselves but are strongly class-conditionally dependent can create a highly informative output. Subsequently, a series of papers show, consistently, that individual accuracy is not good a measure of the joint accuracy of larger feature subsets. The papers provide examples in which the probability of error is estimated for individual variables and for their combinations. The studies find that relatively high individual accuracy can go together with relatively low joint accuracy and vice versa. These results were obtained for class-conditionally independent, binary valued variables (Elashoff et al., 1967; Toussaint, 1971; Cover, 1974); class-conditionally independent, continuous variables (Fang, 1979); and class-conditionally dependent, continuous variables (Cover and Van Campenhout, 1977).

- Barabash (1965) concludes that the amount of information will increase most when information sources are unconditionally independent and class-conditionally dependent. It grows least when the individual information sources are unconditionally dependent and state-conditionally independent. It is additive if information sources are both unconditionally independent and state-conditionally independent.

2.4 Software economics: Does it follow the information goods model?

2.4.1 Software as information good

Economic analyses commonly equate software with code. This way, just like information goods in general, software is considered to be a *public good*: “..each individual’s consumption of such a good leads to no subtraction from any other individual’s consumption of that good..” (Samuelson, 1954). The cost model of information goods, including software, implies that it has high sunk or first copy cost while the cost of each additional copy is small, possibly reaching zero. Consequently, the argument goes, pricing based on the marginal cost of production is bound to end in losses; information must be priced according to the buyer’s value rather than costs.

This economic understanding is very common in economic literature and elsewhere.

For example, Shapiro and Varian (1999) suggest (pp. 20-21): “One of the most fundamental features of information goods is that their cost of production is dominated by the ‘first-copy costs.’ Once the first copy of a book has been printed, the cost of printing another one is only a few dollars. The cost of stamping out an additional CD is less than a dollar, and the vast bulk of the cost of those $80 million movies is incurred prior to the production of the first print.

Or, “Constant fixed costs and zero marginal costs are common assumptions for textbook analysis, but are rarely observed for physical products since there are capacity constraints in nearly every production process. But for information goods, this sort of cost structure is very common-indeed it is the baseline case.” (Varian, 2001)

“..cost based pricing just doesn’t work: a 10 or 20 percent markup on unit cost makes no sense when unit cost is zero. You must price your information goods according to customer value, not according to your production cost.” (Shapiro and Varian, 1999, p. 3)

2.4.2 Software economics

Software economics research is closely related to software engineering. Software economics “seeks to enable significant improvements in software design and engineering through economic reasoning about product, process, and portfolio and policy issues” (Boehm and Sullivan, 2000).

Historically, the principal source of large-scale software was development contracts with the US department of defense (DOD). Ever since those times, software engineering and related literature have typically been limited to the DOD software contract model, despite the major shifts in the software market over the years towards commercial software, mass market, and more (Boehm and Sullivan, 2000; Fayad et al., 2000). In particular, software economics research focused on cost minimization of the code.

In recent years the interest in different perspectives is increasing. Software economics research recently began to emphasize value creation over the cost minimization viewpoint; market characteristics form an important factor under the new perception (Boehm and Sullivan, 2000).

3. DOES HIGHER DATA ACCURACY PRODUCE HIGHER FORECAST ACCURACY?

3.1 INTRODUCTION

Improving data accuracy can be a costly endeavor—it may require re-organization of work processes, acquisition of new technologies (new hardware, new software), and other changes. Therefore there is a critical need to assess the accuracy of the output given whatever use is made of the data, and interpret respective changes in economic terms. The notion "garbage in garbage out" reflects a commonly accepted belief among the MIS community that the accuracy of the output of an information system is tightly and positively linked to the accuracy of its input. However, MIS research has typically focused on contexts in which the output supports only single activities, of an operational, rather than decision-making, nature (Ballou and Tayi, 1999).

Suppose that, in a decision-making context, sales data are used for estimating future demand. Does a lower number of errors in the sales data guarantee that a forecast of future demand will be more accurate ("monotonicity")? If higher data accuracy is not necessarily combined with higher accuracy of a derived forecast, then previously recommended investments in accuracy may not be as valuable as they are believed to be. New approaches may be needed for the identification, assessment, and/or elimination, of undesired consequences of data quality improvements.

This research analyzes the monotonicity assumption acknowledging the diverse use of data in practice, especially in decision-making. Is the relationship between the accuracy of an information system's input and the accuracy of its output monotonic increasing? Is the relationship between the accuracy of an information system's input and the economic value of its output monotonic increasing? This chapter investigates these questions under the condition that the information system uses a single input source for generating the output information. In other words, the information system is viewed as a function that maps the values of a one-dimensional random variable to the values of a one-dimensional random variable (Figure 1.2). It is also assumed that information processing is "noiseless", i.e., the accuracy of the output is determined by the accuracy of the input. Chapter 4, which explores information integration, complements this investigation by scrutinizing conditions on different input sources that imply monotonicity, or non-monotonicity.

The term "more accurate" and a more general term, "more informative", will be defined formally. Afterwards, the formal analysis will refer mostly to the notion of "more informative", however, such analysis applies directly to the special case of "more accurate". The term "has higher information quality" is used in this chapter interchangeably with the two former terms.

The chapter presents a broad condition under which higher accuracy of an information system's input implies higher accuracy of its output—a condition for monotonicity. This condition accounts for both stochastic and deterministic relationships between the variables of interest, rather than the purely deterministic orientation of earlier work. An economic interpretation produces, in addition, a new result on the conditions in which higher input accuracy translates to higher economic value of the output information.

A subsequent part of this chapter characterizes a class of situations in which the proposed theoretical condition is not satisfied, and, in fact, higher input accuracy is translated to lower output accuracy. Current theory misses a contextual factor that can affect accuracy both ways—positive or negative—and therefore should be understood and managed. This factor is *dependence between errors*.

Chapter 3 is organized in the following way: Section 3.2 introduces theoretical foundations. Section 3.3 presents a sufficient condition for monotonicity (the "Monotonicity Theorem"). Section 3.4 extends the formal theory to continuous random variables. Section 3.5 analyzes a linear regression model in which the assumptions do not conform to the sufficient condition for monotonicity in Section 3.4. The effect of measurement errors on the predictive accuracy of the model is derived symbolically, and a simulation that is consistent with scenario #1 in Chapter 1 illustrates the symbolic results. Section 3.6 concludes the chapter.

3.2 THEORETICAL FOUNDATIONS

3.2.1 How information is modeled

Data or information are modeled by an *information structure* (IS). The following examples demonstrate how an IS can model various uses of data. Three different types of uses are considered.

Example 3.2.1.1 (Deterministic setting): Suppose that an information system does not process the data, but rather, outputs the data "as is". For example, consider data on the orders from a product, especially the number of units ordered. Suppose that such data are reported regularly to an employee in charge.

To simplify, it is assumed that the product can be ordered in one of three possible quantities (packages): 100 units, 1000 units, and 10000. Suppose also that the data are not free of errors. Therefore, turning now to the model, the *state set* is *S*={100,1000,10000}; the *signal set* is *Y*={"100", "1000", "10000"}; let the respective IS, denoted ***A***, be the following:

(3.1)

Signal /State	**"100"**	**"1000"**	**"10000"**
100	0.97	0.02	0.01
1000	0.01	0.98	0.01
10000	0.01	0.01	0.98

A shows that, given that a customer orders 100 units, the probability that the reported value will be "100" is 0.97, the probability that the reported value is "1000" is 0.02, and the probability that it is "10000" is 0.01. When 1000 units are actually ordered, the probability that the recorded value is "1000" is 0.98, the probability that it is "100" is 0.01, the probability that it is "10000" is 0.01, and so on. If the data were error-free, the respective IS would be an identity matrix, i.e., a square matrix whose diagonal elements are 1s and whose off-diagonal elements are all 0s.◘

Example 3.2.1.2 (Processing): In many cases data undergo processing. For example, the employee mentioned earlier may want to be informed about revenues instead of volumes of sales, such that sales data may have to be processed.

Suppose that revenue is calculated based on data about the number of units ordered, and the price of each package is fixed such that the revenue per package can have one of three possible values, e.g., 400, 3600, and 32000. The state set is therefore S={400,3600,32000}, and the signal set is Y={"400","3600","32000"}. Due to the fact that the output of information processing forms, in this instance, a one-to-one function from the sales value set to the revenue value set, any errors in the input are propagated directly to the output. Therefore, matrix 3.2 is the same as matrix 3.1:

(3.2)

Signal /State	**"400"**	**"3600"**	**"32000"**
400	0.97	0.02	0.01
3600	0.01	0.98	0.01
32000	0.01	0.01	0.98

The formal analysis in this study assumes that information processing amounts to a one-to-one function. This assumption serves the selected approach, which assesses the *maximal* accuracy that can be achieved in using a given input. (Lemma 4.2.1.1 in the next chapter proves that a one-to-one function is optimal in this sense when information processing combines multiple inputs. That lemma can be used to demonstrate a similar property when a single input is involved.)◘

Example 3.2.1.3 (Uncertainty): Consider, for example, a situation in which data on the number of units ordered help in predicting the size of a subsequent order. (For example, the estimate may simply apply the rule that a customer's order is identical to whatever he/she ordered last.) Therefore, the state set may be S={100^F, 1000^F, 10000^F}—the superscript has been added to differentiate this state set from the previous state set. The signal set may remain as before *Y*={"100", "1000", "10000"}. The IS below (3.3), denoted ***B,*** is different from ***A***. The matrix reflects, in addition to recording errors, an element of uncertainty due to the fact that a customer may change her or his order from time to time. To simplify, this IS assumes, for example, that customers do not skip an order or stop ordering.

(3.3)

Signal /State	**"100"**	**"1000"**	**"10000"**
100^F	0.874	0.116	0.01
1000^F	0.020	0.941	0.039
10000^F	0.023	0.055	0.922

To clarify the compound nature of the errors that ***B*** models, note that it is the product of matrix 3.4, and ***A*** (the numbers were rounded). The Markov matrix under 3.4 mirrors a stochastic relationship between the number of units ordered and a subsequent order, and ***A*** portrays recording errors.

(3.4)

State/ State	**100**	**1000**	**10000**
100^F	0.9	0.1	0
1000^F	0.01	0.96	0.03
10000^F	0.014	0.046	0.94

◘

3.2.2 Evaluating and comparing information sources

Accuracy and informativeness are operationalized using statistical terms—Blackwell's sufficiency criterion. These definitions are similar, for example, to the definitions in (Barua et al., 1989). I first provide these definitions and then discuss Blackwell's Theorem, which asserts the equivalence of higher accuracy / informativeness and economic superiority.

Definition (3.2.2.1): *Information structure (IS).* Let S denote a finite set of *states of the world*, and Y a finite set of *signals*. The elements of each of these sets are mutually exclusive and exhaustive, that is, one and only one state will occur, and one and only one signal will be observed. An *information structure* $\boldsymbol{A}$:Sx$Y \rightarrow [0,1]$ is described by a Markov matrix where the value in row *i*, column *j*, shows the probability that a signal $y_j \in Y$ will be produced given that a state $s_i \in S$ occurs.◙

A ranking of ISs by their accuracy, or informativeness, is defined using Blackwell's sufficiency criterion (Definition 2.1.3):

Definition (3.2.2.2): *Informativeness.* Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be two ISs defined on SxY_{A} and SxY_{B} respectively. $\boldsymbol{A}$ is *more informative*, or *generally more informative* (GMI), than $\boldsymbol{B}$ if there exists a Markov matrix $\boldsymbol{M}$ such that $\boldsymbol{AM}$=$\boldsymbol{B}$.◙

Definition (3.2.2.3): *Accuracy.* Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be two ISs defined on SxY_{A} and SxY_{B}, respectively. For every $s_i \neq s_k$ in S there exists $y^A \in Y_{\mathrm{A}}$ such that $\boldsymbol{A}(s_i,y^A) \neq \boldsymbol{A}(s_k,y^A)$. Similarly, for every $s_i \neq s_k$ in S there exists $y^B \in Y_B$ such that $\boldsymbol{B}(s_i, y^B) \neq \boldsymbol{B}(s_k, y^B)$. Then, $\boldsymbol{A}$ is *more accurate* than $\boldsymbol{B}$ if there exists a Markov matrix $\boldsymbol{M}$ such that $\boldsymbol{AM}=\boldsymbol{B}$.◙

"Accuracy" and "informativeness" are closely related notions. "More informative", however, is a broader term. It does not require that $\boldsymbol{A}$ and $\boldsymbol{B}$ distinguish between states, i.e., that the conditional probability function given each state is unique. The relaxation in the definition of "more informative" permits a situation in which the set of states that $\boldsymbol{B}$ distinguishes is a subset of the set of states that $\boldsymbol{A}$ distinguishes, i.e., $\boldsymbol{A}$ has higher resolution—also known as more precise, or finer—than $\boldsymbol{B}$. (See Marschak and Radner, 1972.)

<u>Example 3.2.2.1:</u> $\boldsymbol{A}$ (IS 3.1) is more accurate then the IS below (3.5), denoted $\boldsymbol{C}$.

(3.5)

Signal /State	**"100*"**	**"1000*"**	**"10000*"**
100	0.95	0.03	0.02
1000	0.04	0.94	0.02
10000	0.03	0.02	0.95

$\boldsymbol{C}$ can be obtained from $\boldsymbol{A}$ by multiplying it by the Markov matrix in 3.6, denoted $\boldsymbol{M}$; $\boldsymbol{AM}=\boldsymbol{C}$. Therefore, Definition 3.2.2.3 is satisfied:[1]

[1] The numbers in $\boldsymbol{C}$ (3.5) were rounded.

(3.6)

Signal /Signal	"100*"	"1000*"	"10000*"
"100"	0.98	0.01	0.01
"100"	0.03	0.96	0.01
"100"	0.02	0.01	0.97

◘

Definition (3.2.2.4): Information structure (IS) Model. Let S denote a finite set of *states of the world*, Y a finite set of *signals*, and D the decision-maker's finite set of feasible *actions*. The elements of each of these sets are assumed to be mutually exclusive and exhaustive.

Let $\boldsymbol{p}$ be a vector of *a priori* probabilities of the states of the world, $\boldsymbol{A}$ an IS as defined above, $\boldsymbol{R}$:YxD $\rightarrow$[0,1], describing the decision-maker's *decision rule*, is given by a Markov matrix in which the value in row *i* column *j* is the probability that an action $d_j \in D$ will be adopted, given that the observed signal is $y_i \in Y$.

Finally, $\boldsymbol{U}$ is the *payoff* function $\boldsymbol{U}$:Dx$S \rightarrow \Re$, associating a payoff to each pair of strategy and state of nature. $\boldsymbol{U}_{i,j}$ is the payoff if an action $d_i \in D$ is adopted when the actual state is $s_j \in S$.

The expected payoff given $\boldsymbol{p}$, $\boldsymbol{A}$, $\boldsymbol{R}$, $\boldsymbol{U}$, is tr($\boldsymbol{ARUp'}$)=$\Sigma_{s,y,d}$ $\boldsymbol{U}(d,s)\boldsymbol{p}(s)\boldsymbol{A}(s,y)\boldsymbol{R}(y,d)$. "tr" refers to the mathematical trace operator, and $\boldsymbol{p'}$ stands for a square matrix containing the elements of $\boldsymbol{p}$ in its main diagonal. ◙

A decision maker is assumed to choose a decision rule that maximizes her expected payoff, i.e., Max $_R$ tr($\boldsymbol{ARUp'}$).

Blackwell's Theorem (3.2.2.1). Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be two ISs defined on $S\text{x}Y_A$ and $S\text{x}Y_B$, respectively.

$\boldsymbol{A}$ is GMI/more accurate than $\boldsymbol{B}$ if and only if for any set of actions D, payoff matrix $\boldsymbol{U}$, and vector $\boldsymbol{p}$ of a priori probabilities, $\text{Max}_{R_B}\text{tr}(\boldsymbol{BR_BUp'}) \leq \text{Max}_{R_A}\text{tr}(\boldsymbol{AR_AUp'})$, where $\boldsymbol{R_A}$, $\boldsymbol{R_B}$, denote the decision rule defined on $Y_A\text{x}D$, $Y_B\text{x}D$, respectively.◙

<u>Example 3.2.2.2:</u> $\boldsymbol{A}$ (IS 3.1) is more accurate than $\boldsymbol{C}$ (IS 3.5), therefore, according to Blackwell's Thoerem, a decision-maker should have at least as high maximal expected payoff when using $\boldsymbol{A}$ as when using $\boldsymbol{C}$, no-matter her payoff function, action set, and prior probability distribution. To further illustrate this understanding, suppose that $\boldsymbol{p}$=(0.1,0.6,0.3), $D=\{d_1,d_2,d_3,d_4\}$, and the payoff matrix $\boldsymbol{U}$ is described by the matrix:

(3.7)

State /Action	100	1000	10000
d_1	100	-5	60
d_2	75	15	65
d_3	30	90	20
d_4	50	40	70

Calculations will show that the best payoff using $\boldsymbol{A}$ is 83.76. This payoff can be achieved using the decision rule matrix under 3.8:

(3.8)

Action /Signal	d_1	d_2	d_3	d_4
"100"	1	0	0	0
"1000"	0	0	1	0
"10000"	0	0	0	1

The best payoff using ***C*** is 81.42. It can be achieved using the same decision rule (3.8). In agreement with theory, ***A*** generates higher payoff.

◘

3.3 WHEN IS ACCURACY GUARANTEED TO CHANGE CONSISTENTLY ACROSS DIFFERENT APPLICATIONS?

The IS theory will now be applied to address the main issue of this chapter.

3.3.1 IS that informs about a given state using a given signal

First, I propose a formal definition of the situation in which a signal is used to inform about a state.

Definition (3.3.1.1): ***A*** *informs about state* ***S*** *using signal* ***Y***. Let ***Y***, ***S***, denote random variables, taking values in the finite sets *Y*, *S*, respectively. If ***A***=[Pr(***Y***=*y*|***S***=*s*)], then ***A*** *informs about* ***S*** *using* ***Y***. ◙

By definition, ***A*** is an IS.

Definition 3.3.1.1 is easily generalizable to situations in which a signal is used to inform about $n \geq 1$ states. The next examples illustrate Definition 3.3.1.1.

Example 3.3.1.1: (ISs that inform about two or more given states using a given signal): ***A*** (IS 3.1) models the information provided by a data source on the number of units ordered from a chosen product. IS 3.2 models the information provided by the information system on respective revenues when the system uses the given data source. ***B*** (3.3) models the information provided by the information system on subsequent orders from the product, when such forecast is based on the former data.

It can be shown that the above ISs inform about three different states using the same signal. Take, in particular, ***A*** and ***B***. Let ***Y*** denote a random variable that models the order data, ***S*** a random variable that models the actual order, and ***S'*** a random variable that models the next order. Furthermore, let s_1 denote the state 100; s_2, the state 1000; s_3, the state 10000; let s'_1 denote the state 100^F; s'_2, the state 1000^F; s'_3, the state 10000^F; let y_1 denote the signal "100"; y_2 - "1000"; y_3 - "10000". Then, the joint probability function of ***Y***, ***S***, ***S'***, is:

(3.9)

$Pr(y_1,s_1,s'_1)$ =0.0873	$Pr(y_2,s_1,s'_1)$ =0.0018	$Pr(y_3,s_1,s'_1)$ =.0009	$Pr(y_1,s_2,s'_1)$ =.00006	$Pr(y_2,s_2,s'_1)$ =.00588	$Pr(y_3,s_2,s'_1)$ =.00006	$Pr(y_1,s_3,s'_1)$ =.0000405	$Pr(y_2,s_3,s'_1)$ =.0000405	$Pr(y_3,s_3,s'_1)$ =.003969
$Pr(y_1,s_1,s'_2)$ =.0097	$Pr(y_2,s_1,s'_2)$ =.0002	$Pr(y_3,s_1,s'_2)$ =.0001	$Pr(y_1,s_2,s'_2)$ =.00576	$Pr(y_2,s_2,s'_2)$ =.56448	$Pr(y_3,s_2,s'_2)$ =.00576	$Pr(y_1,s_3,s'_2)$ =.0001395	$Pr(y_2,s_3,s'_2)$ =.0001395	$Pr(y_3,s_3,s'_2)$ =.013671
$Pr(y_1,s_1,s'_3)$ =0	$Pr(y_2,s_1,s'_3)$ =0	$Pr(y_3,s_1,s'_3)$ =0	$Pr(y_1,s_2,s'_3)$ =.00018	$Pr(y_2,s_2,s'_3)$ =.01764	$Pr(y_3,s_2,s'_3)$ =.00018	$Pr(y_1,s_3,s'_3)$ =.00282	$Pr(y_2,s_3,s'_3)$ =.00282	$Pr(y_3,s_3,s'_3)$ =.27636

The consistency of ***A*** and ***B*** with this joint probability function can be verified, so Definition 3.3.1.1 is satisfied. ◘

Example 3.3.1.2 (ISs that inform about two different states using different signals): Suppose that the joint probability function of some ***Y***, ***S***, and ***S'***, is as given by 3.10, and the joint probability function of ***Y'***, ***S***, and ***S'***, is as given by 3.11:

(3.10)

$Pr(y_1,s_1,s'_1)$ = .05	$Pr(y_1,s_1,s'_2)$ =.3	$Pr(y_1,s_2,s'_1)$ =.05	$Pr(y_1,s_2,s'_2)$ =.05	$Pr(y_2,s_1,s'_1)$ =.05	$Pr(y_2,s_1,s'_2)$ =.05	$Pr(y_2,s_2,s'_1)$ =.35	$Pr(y_2,s_2,s'_2)$ =.1

(The rest of the probabilities are equal zero.)

(3.11)

$Pr(y'_1,s_1,s'_1)$ = .05	$Pr(y'_1,s_1,s'_2)$ =.2	$Pr(y'_1,s_2,s'_1)$ =.2	$Pr(y'_1,s_2,s'_2)$ =.1	$Pr(y'_2,s_1,s'_1)$ =.05	$Pr(y'_2,s_1,s'_2)$ =.15	$Pr(y'_2,s_2,s'_1)$ =.2
$Pr(y'_2,s_2,s'_2)$ = .05						

(The rest of the probabilities are equal zero.)

Note the joint probability functions described by 3.10 and 3.11 satisfy a necessary condition, i.e., they imply the same two-dimensional marginal probability distribution for ***S***, ***S'***, described by the probability function under 3.12:

(3.12)

$Pr(s_1,s'_1)$ =.1	$Pr(s_1,s'_2)$ =.35	$Pr(s_2,s'_1)$ =.4	$Pr(s_2,s'_2)$ =.15

(The rest of the probabilities are equal zero.)

Based on 3.10, ***D***=[Pr(***Y***=y |***S***=s)] is:

(3.13)

Signal /State	y_1	y_2
s_1	.2	.8
s_2	.7	.3

Similarly, based on 3.11, $\boldsymbol{D'}$=[Pr($\boldsymbol{Y'}$=y' |$\boldsymbol{S'}$=s')] is:

(3.14)

Signal /State	y'_1	y'_2
s'_1	.56	.44
s'_2	.55	.45

Each of ***D*** and ***D'*** is derived from a different joint probability function. It can be verified that ***D*** could not have been derived from the joint probability function on which ***D'*** is based, and vice versa.

◘

3.3.2 Monotonicity

Definition 3.3.1.1 is applied in a definition that formalizes the notion of monotonicity.

Definition (3.3.2.1): *Monotonicity*. ***A*** and ***A'*** inform about ***S*** and ***S'***, respectively, using $\boldsymbol{Y}^{A}$. ***B*** and ***B'*** inform about ***S*** and ***S'***, respectively, using $\boldsymbol{Y}^{B}$. If ***B*** is GMI than ***A*** implies ***B'*** is GMI than ***A'*** then the relationship between the informativeness of ***B***, ***A***, and the informativeness of ***B'***, ***A'***, is *monotonic* ◙

When a relationship as specified in Definition 3.3.2.1 is monotonic, a shift from using $\boldsymbol{Y}^{\mathbf{A}}$ to using $\boldsymbol{Y}^{\mathbf{B}}$ increases the quality of the information about $\boldsymbol{S}$ as well as $\boldsymbol{S'}$.

Next, the Monotonicity Theorem (3.3.2.1) specifies a condition such that Definition 3.3.2.1 is satisfied. A major premise of this theorem is that the signals do not provide additional information on $\boldsymbol{S'}$ beyond the information supplied by $\boldsymbol{S}$. Section 3.5 constructs an example that clarifies this condition. The state-conditional independence that the Monotonicity Theorem requires is revealed there as independence between errors, i.e., between the *stochastic disturbance* and the *measurement error*.

The theorem assumes two state variables, but it can be directly generalized to $n \geq 2$ variables.

Lemma 3.3.2.1: $\boldsymbol{A}$ and $\boldsymbol{A'}$ are ISs that inform about $\boldsymbol{S}$, $\boldsymbol{S'}$, respectively, using $\boldsymbol{Y}$. If $\boldsymbol{Y}$ is independent of $\boldsymbol{S'}$ given $\boldsymbol{S}=s$ for every $s \in S$, then $\boldsymbol{A'}=\boldsymbol{QA}$, $\boldsymbol{Q}=[\Pr(\boldsymbol{S}=s|\boldsymbol{S'}=s')]$.

Proof:

Since $\boldsymbol{A}$ and $\boldsymbol{A'}$ inform about $\boldsymbol{S}$, $\boldsymbol{S'}$, respectively, using $\boldsymbol{Y}$, the elements of $\boldsymbol{A'}$ are of the form:

$\Pr(\boldsymbol{Y}=y\,|\boldsymbol{S'}=s')=\Sigma_s \Pr(\boldsymbol{Y}=y,\boldsymbol{S}=s,\boldsymbol{S'}=s') \,/\, \Sigma_y \Sigma_s \Pr(\boldsymbol{Y}=y,\boldsymbol{S}=s,\boldsymbol{S'}=s')$

$= \Sigma_s \Pr(\boldsymbol{Y}=y,\boldsymbol{S'}=s'|\boldsymbol{S}=s)\cdot\Pr(\boldsymbol{S}=s) \,/\, \Pr(\boldsymbol{S'}=s')$

By assumption, $\boldsymbol{Y}$ is independent of $\boldsymbol{S'}$ given $\boldsymbol{S}=s$ for every $s \in S$. Therefore, the above can be re-expressed:

$\Pr(\boldsymbol{Y}=y\ |\boldsymbol{S'}=s')=\Sigma_s \Pr(\boldsymbol{Y}=y|\boldsymbol{S}=s)\cdot\Pr(\boldsymbol{S'}=s'|\boldsymbol{S}=s)\cdot\Pr(\boldsymbol{S}=s)\ /\ \Pr(\boldsymbol{S'}=s')=$

$\Sigma_s \Pr(\boldsymbol{Y}=y|\boldsymbol{S}=s)\cdot\Pr(\boldsymbol{S}=s|\boldsymbol{S'}=s')$

Or, using a different notation:

$\boldsymbol{A'}=\boldsymbol{QA}$, where $\boldsymbol{Q}=[\Pr(\boldsymbol{S}=s|\ \boldsymbol{S'}=s')]$.

◙

Theorem 3.3.2.1 (Monotonicity) $\boldsymbol{A}$, $\boldsymbol{A'}$, are ISs that inform about $\boldsymbol{S}$, $\boldsymbol{S'}$, using $\boldsymbol{Y}^A$. Similarly, $\boldsymbol{B}$, $\boldsymbol{B'}$, are ISs that inform about $\boldsymbol{S}$, $\boldsymbol{S'}$, using $\boldsymbol{Y}^B$. If $\boldsymbol{Y}^{\mathbf{A}}$ is conditionally independent of $\boldsymbol{S'}$ given $\boldsymbol{S}=s$ for every $s\in S$, and $\boldsymbol{Y}^{\mathbf{B}}$ is conditionally independent of $\boldsymbol{S'}$ given $\boldsymbol{S}=s$ for every $s\in S$, then the relationship between the informativeness of $\boldsymbol{B}$, $\boldsymbol{A}$, and the informativeness of $\boldsymbol{B'}$, $\boldsymbol{A'}$, is monotonic.

Proof:

According to Lemma 3.3.1:

(1) $\boldsymbol{A'}=\boldsymbol{QA}$, $\boldsymbol{B'}=\boldsymbol{QB}$, where $\boldsymbol{Q}=[\Pr(\boldsymbol{S}=s|\boldsymbol{S'}=s')]$.

Suppose that $\boldsymbol{B}$ is GMI than $\boldsymbol{A}$, then by definition there exists a Markov matrix $\boldsymbol{N}$ such that $\boldsymbol{BN}=\boldsymbol{A}$.

$\Rightarrow \boldsymbol{A'}=\boldsymbol{QA}=\boldsymbol{QBN}$

(1) Implies that $\boldsymbol{QBN}=\boldsymbol{B'N}$

$\Rightarrow \boldsymbol{A'}=\boldsymbol{QA}=\boldsymbol{QBN}=\boldsymbol{B'N}$

N is a Markov Matrix, therefore, by definition, ***B'*** is GMI than ***A'***.

◙

The independence condition of the Monotonicity Theorem is satisfied when ***S'*** is a deterministic function of ***S***. In other words, the condition of the theorem is met when the relationship between the phenomenon designated by the input and the phenomenon designated by the output is deterministic, rather than stochastic. This corollary applies to many practical problems (involving, for example, certain financial management formulas, operations management formulas, physics laws, geometry, and other domains). Moreover, it coincides with current data quality theory, which, as indicated earlier, has a deterministic standpoint. The Monotonicity Theorem, therefore, establishes a generalization of the current theory.

Based on Blackwell's Theorem, the Monotonicity Theorem can also be interpreted from an economic perspective. Viewed from this angle, the Monotonicity Theorem specifies a sufficient condition—the independence condition as above—such that higher informativeness (and maximal expected utility) of the input of an information system translates to higher maximal expected utility of the output of this system.

3.4 EXTENSION OF THE THEORY TO CONTINUOUS RANDOM VARIABLES

The formal theory that was developed in Section 3.3 for random variables that can take only a finite number of different values is extended in this section to continuous random variables.

Definition (3.4.1): *Information structure (IS).* Let S denote a set of *states of the world*, and Y a set of *signals*. The elements of each of these sets are mutually exclusive and exhaustive, that is, one and only one state will occur, and one and only one signal will be produced by the information system. A function f:Sx$Y \rightarrow \Re^+$ such that for any $s \in S$, $f(y|s)$ is a probability density function, is an *information structure*.◙

Definition (3.4.2): *More Informative* Let f^A and f^B be two ISs defined on SxY_A and SxY_B, respectively. f^A is *more informative*, or *generally more informative* (GMI) than f^B if there exists a function T:Y_Bx$Y_{\boldsymbol{A}} \rightarrow \Re^+$ such that $f^B(y^B|s)=\int T(y^B, y^A)\cdot f^A(y^A|s)\, dy^A \;\; \forall\, s \in S$, $y^B \in Y_{\boldsymbol{B}}$, where $0 < \int T(y^B, y^A)dy^A < \infty \;\; \forall y^B \in Y_{\boldsymbol{B}}$, and $\int T(y^B, y^A)dy^B = 1 \;\; \forall\, y^A \in Y_{\boldsymbol{A}}$.◙

Blackwell's theorem applies for any state and signal sets (Blackwell, 1953), (Boll, 1955).

Definition (3.4.3): *$\boldsymbol{f}$ informs about state $\boldsymbol{S}$ using signal $\boldsymbol{Y}$* Let $\boldsymbol{Y}$, $\boldsymbol{S}$, denote continuous random variables, taking values in the sets Y, and S respectively. If for every $s \in S$, a function f:Sx$Y \rightarrow \Re^+$ is the conditional density of $\boldsymbol{Y}$ given $\boldsymbol{S}$=s, then f *informs about $\boldsymbol{S}$ using $\boldsymbol{Y}$* .◙

It can be easily seen that such f is an IS.

Definition (3.4.4): *Monotonicity.* f^A, $f^{A'}$, are ISs that inform about $\boldsymbol{S}$, $\boldsymbol{S'}$, respectively, using $\boldsymbol{Y}^A$. Similarly, f^B, $f^{B'}$, are ISs that inform about $\boldsymbol{S}$, $\boldsymbol{S'}$, respectively, using $\boldsymbol{Y}^B$. If f^B is

GMI than f^{A} implies $f^{B'}$ is GMI than $f^{A'}$ then the relationship between the informativeness of f^{B}, f^{A}, and the informativeness of $f^{B'}$, $f^{A'}$, is *monotonic*.◙

Lemma 3.4.1: f, f', are ISs that inform about $\boldsymbol{S}$, $\boldsymbol{S'}$, respectively, using $\boldsymbol{Y}$. If $\boldsymbol{Y}$ is conditionally independent of $\boldsymbol{S'}$ given $\boldsymbol{S}=s$ for every $s \in S$, then $f'(y|s') = \int f(y|s) \cdot g(s|s') ds$, $y \in Y$, $s' \in S'$, where $g(s|s')$ denotes the conditional density of $\boldsymbol{S}$ given $\boldsymbol{S'}=s'$.

Proof:

Let $f_{YS'}(y,s')$ denote the joint density function of $\boldsymbol{Y}$, $\boldsymbol{S'}$; $f_{YS'S}(y,s',s)$ - the joint density function of $\boldsymbol{Y}$, $\boldsymbol{S'}$, $\boldsymbol{S}$; $f_{YS'|S}(y,s'|s)$ - the conditional density function of $\boldsymbol{Y}$, $\boldsymbol{S'}$, given $\boldsymbol{S}=s$; and $f_{S}(s)$, $f_{S'}(s')$ - the probability density functions of $\boldsymbol{S'}$, $\boldsymbol{S}$, respectively.

$$f'(y|s') = f_{YS'}(y,s') \,/\, f_{S'}(s') = [\int f_{YS'S}(y,s',s)\, ds \,] \,/\, f_{S'}(s') =$$

$$= [\int f_{YS'|S}(y,s'|s) \cdot f_{S}(s)\, ds \,] \,/\, f_{S'}(s') = \int f_{YS'|S}(y,s'|s) \cdot f_{S}(s) \,/\, f_{S'}(s')\, ds$$

By assumption, $\boldsymbol{Y}$ is conditionally independent of $\boldsymbol{S'}$ given $\boldsymbol{S}=s$ for every $s \in S$. Therefore, the above can be re-expressed:

$$f'(y|s') = \int f(y|s) \cdot f_{S'|S}(s'|s) \cdot f_{S}(s) \,/\, f_{S'}(s')\, ds = \int f(y|s) \cdot g(s|s')\, ds$$

.◙

Theorem 3.4.1 (Monotonicity) f^{A}, $f^{A'}$, are ISs that inform about $\boldsymbol{S}$, $\boldsymbol{S'}$, respectively, using $\boldsymbol{Y}^{A}$. Similarly, f^{B}, $f^{B'}$, are ISs that inform about $\boldsymbol{S}$, $\boldsymbol{S'}$, respectively, using $\boldsymbol{Y}^{B}$. If $\boldsymbol{Y}^{\boldsymbol{A}}$ is conditionally independent of ***S'*** given ***S***=*s* for every $s \in S$, and $\boldsymbol{Y}^{\mathbf{B}}$ is conditionally independent of ***S'*** given ***S***=*s* for every $s \in S$, then the relationship between the informativeness of f^{B}, f^{A}, and the informativeness of $f^{B'}$, $f^{A'}$, is monotonic.

Proof:

Following the assumption of independence in this theorem, Lemma 3.4.1 implies that:

(1) $f^{A'}(y^{A}|s')=\int f^{A}(y^{A}|s)\cdot g(s|s')ds$

(2) $f^{B'}(y^{B}|s')=\int f^{B}(y^{B}|s)\cdot g(s|s')ds$

Suppose that f^{B} is GMI than f^{A}, then, by definition

$f^{A}(y^{A}|s)=\int T(y^{B}, y^{A})\cdot f^{B}(y^{B}|s)dy^{B}, \quad \forall\, y^{A}, s$

such that $T: Y^{B} \times Y^{A} \rightarrow \Re^{+},\ 0<\int T(y^{B}, y^{A})dy^{A}<\infty\ \ \forall y^{B},\ \ \int T(y^{B}, y^{A})dy^{B}=1\ \ \forall\, y^{A}$.

Therefore:

$f^{A'}(y^{A}|s')= \int[\int T(y^{B}, y^{A})\cdot f^{B}(y^{B}|s)\, dy^{B}\ ]\cdot g(s|s')ds$

$=\iint T(y^{B},y^{A})\cdot f^{B}(y^{B}|s)\cdot g(s|s')dy^{B}ds =$

$=\iint T(y^{B},y^{A})\cdot f^{B}(y^{B}|s)\cdot g(s|s')\, ds dy^{B}=\int T(y^{B},y^{A})\cdot f^{B'}(y^{B}|s')dy^{B}$

Therefore, by definition, $f^{B'}(y^{B}|s')$ is GMI than $f^{A'}(y^{A}|s')$.

.◙

3.5 NON-MONOTONICITY

In this section I provide a concrete characterization of a class of situations in which the assumptions violate the independence condition of the Monotonicity Theorem, and, correspondingly, higher accuracy does not propagate in the desired way.

The present characterization concentrates on a type of *non-classical measurement errors*, in the context of a simple linear regression model, i.e., data that have errors of that type are used for the prediction of a value of interest through a known linear regression model. In agreement with the scope of this chapter, the linear regression model has only one independent variable.

3.5.1 A counter example

3.5.1.1 The scenario

Consider the model:

$$y^* = \beta x^* + \varepsilon$$

This model represents a true relationship between x^* and y^*. x^* is a random variable, ε—the "disturbance"—is a random variable, $\varepsilon \sim N(0,\sigma_\varepsilon^2)$, and is independent of x^*. The value of β is known. (This way, results pertaining to accuracy can be attributed only to the inherent uncertainty in the model as demonstrated by ε, and/or to characteristics of the input to the model. This restriction is, again, consistent with the scope of the inquiry—see in Section 3.1.)

Suppose that x^* is unobserved, and an observed variable, x, is such that:

$$x = x^* + u$$

u, the error term, is a random variable, $u \sim N(0,\sigma_u^2)$, and is independent of x^*. However, u and ε are not independent, such that their covariance is not equal zero, i.e., $\sigma_{u\,\varepsilon} \neq 0$.

Since x^* is not available, x is used both as an estimate of x^*, and as the sole input to the calculation of ŷ, the predicted value of y^*:

$$\hat{y} = \beta x = \beta(x^* + u)$$

Next, turn to the accuracy of prediction in this example.

The prediction error is:

$$y^*\text{-}\hat{y} = \beta x^* + \varepsilon - \beta(x^* + u) = \varepsilon - \beta\, u$$

The mean of the prediction error is zero:

$$\mathrm{E}[y^*\text{-}\hat{y}]=\mathrm{E}[\varepsilon - \beta\, u]=0$$

And the variance of the prediction error is given by:

$$\mathrm{Var}[y^*\text{-}\hat{y}]= \mathrm{Var}[\varepsilon - \beta\, u]= \sigma_\varepsilon^{2}+\beta^2\sigma_u^{2}- 2\beta\sigma_{u\,\varepsilon}$$

3.5.1.2 Scenario violates the independence condition in the Monotonicity Theorem

The value of x can be viewed as a signal that is used both for informing about the value of x^*, and for informing about the value of y^*.

Given a value of x^*, is x independent of y^*? The answer to this question is negative. Here is a why:

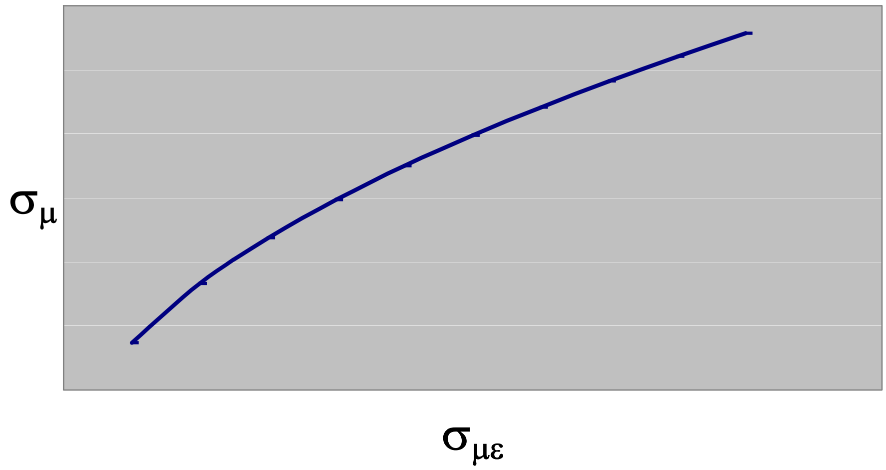
Given a value of x^*, x has the same distribution as u. Similarly, given a value of x^*, y^* has the same distribution as ε. However, u, by assumption, is not independent of ε, and furthermore, the conditions in this scenario imply that unconditional independence overlaps with conditional independence given any value of x^* (see for example Barabash, 1965, or Hussain, 1972). Therefore, given a value of x^*, x and y^* are not independent.

Due to the suggested dependence, the Monotonicity Theorem does not apply. In other words, if instead of x another variable is used to inform about x^*, such that informativeness increases due to this exchange, the theorem does not guarantee that informativeness will also increase when the latter variable is used for informing about y^*.

3.5.1.3 No monotonicity

The calculation of Var[y^*-ŷ] —the variance of the prediction error—shows that $\sigma_{u\varepsilon}$ is one of the determinants of this variance. When $\sigma_{u\varepsilon}$=0, i.e., ε, u, are independent of each other, a lower variance of the measurement error (σ_u^2) is translated to a lower variance of the prediction error. However, suppose that $\beta > 0$, and some change in the data that lowers σ_u^2 decreases $\sigma_{u\varepsilon}$ at the same time. If, consequently, $\sigma_{u\varepsilon}$<0 is low enough, then Var[y^*-ŷ] will grow higher despite a lower σ_u^2. Alternatively, when $\beta < 0$, if a lower σ_u^2 goes together with higher $\sigma_{u\varepsilon}$, i.e., if $\sigma_{u\varepsilon}$>0 is high enough, then, again, Var[y^*-ŷ] will be higher despite the lower σ_u^2.

FIGURE 3.1: Prediction error variance isocurve $\sigma_\varepsilon^2+\beta^2\sigma_u^2 - 2\beta\sigma_{u\varepsilon} = c$



It is easy to see that a suitable change in $\sigma_{u\varepsilon}$ can also contribute to a lower, instead of higher, variance of the prediction error.

Figure 3.1 shows a curve on which the variance of the prediction error is constant. An increase in σ_u^2, the variance of the measurement error, will position Var[y*-ŷ] above the curve, but a simultaneous increase in the covariance $\sigma_{u\varepsilon}$ can bring Var[y*-ŷ] back on the curve (and even drive it below the curve). In other words, Var[y*-ŷ] can remain the same, and may even decrease, despite an increase in σ_u^2. (Figure 3.1 reflects an assumption that β>0.)

Error variance is a measure of accuracy in the scenario in Section 3.5.1.1. It can be shown that it is also consistent with the ranking by accuracy/ informativeness that accompanies the IS model. The analysis conducted by Clemen and Winkler (1985) indicates, among the rest, that the ranking based on Blackwell's sufficiency criterion is consistent with the rankings that the measurement error variance and prediction error variance generate. Therefore, if a lower σ_u^2 is accompanied by a higher Var[y*-ŷ], it implies that Definition 3.4.4 (Monotonicity) is not satisfied, i.e., there is no monotonicity.

To conclude, the sufficient condition of the Monotonicity Theorem is satisfied when u and ε are independent. In this case, as predicted by the theorem, higher accuracy of the data has positive influence on the prediction accuracy of the regression model. However, the sufficient condition of the Monotonicity Theorem is violated when the dependence between u and ε enters as an additional factor. Under such circumstances a change in the accuracy of the input data can affect the results either direction—positive or negative.

A series of simulations illustrates these analytical results. The simulations were performed with GAUSS Light, a mathematical /statistical analysis environment.

3.5.2 Simulation

The example corresponds to scenario #1 in Chapter 1. The assumptions on the distributions of the disturbance and measurement errors have, for the most part, been relaxed, such that they agree with the assumptions of the scenario in 3.5.1.1.

A linear regression model that matches the survey output in scenario #1 is:

$$d^*_2 = 0.5d^*_1 + \varepsilon$$

Where d^*_1 , d^*_2 , denote the number of equipment units in use by the customer, and the demand for the new component during the first year, respectively.

Each series of simulations first generated 500 data items, d^*_{1i}, i=1,2,..,500, from a uniform distribution $d^*_1 \sim U[85,115]$. Next, 500 items corresponding to the stochastic disturbance ε were produced, $\varepsilon \sim N(0,9)$. Subsequently, the values of d^*_{2i} i=1,2,..,500, were derived according to the linear model above.

In each simulation a certain percentage of the synthesized data set for d^*_1 was modified, through instances of measurement errors u_i, i=1,2,..,500, which were created for that purpose. Three parameters of the measurement error were manipulated: (1) the error rate, i.e., the percentage of the data contaminated with measurement error, (2) the variance of the error ($u \sim N(0,4)$, $u \sim N(0,9)$, $u \sim N(0,18)$), and (3) the correlation between u and ε.

Observed values, d_{1i} , i=1,2,..,500, were calculated such that:

$$d_{1i} = d^*_{1i} + u_i$$

The values of d_{1i} served to create estimates $\check{d}_{2i}$ of d^*_{2i} :

$$\check{d}_{2i} = 0.5 d_{1i}, \quad i=1,2,\ldots,500$$

Prediction accuracy was evaluated using two popular measures called Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE) (Greene, 2000). MAPE is defined by:

$$\text{MAPE} = 1/n \ \Sigma_i |y^*_i - \hat{y}_i| / \ y^*_i$$

and in this case:

$$\text{MAPE} = 1/500 \ \Sigma_i |d^*_{2i} - \check{d}_{2i}| / d^*_{2i}$$

RMSE is defined by:

$$\text{RMSE} = \sqrt{1/n\Sigma_i(y^*_i\text{-}\hat{y}_i)^2}$$

and in this case:

$$\text{RMSE} = \sqrt{1/500\Sigma_i(d^*_{2i} - \check{d}_{2i})^2}$$

The empirical results—such as captured by Figure 3.2 and Figure 3.3—illustrate the theoretical conclusions. Each of the graphs summarizes the results of 15 experiments in a different way. In these experiments, the percentage of measurement errors varies from a minimum of 5% to a maximum of 25% of the data, u~$N(0,9)$, and different correlation coefficients vary, and are grouped under three categories, [-0.2,+0.2],[+0.6,+0.8],[-0.5,-0.7]. As the graphs show, positive correlation between u and ε improves the accuracy of prediction, and negative correlation between u and ε degrades it. In addition, when correlation is negative, an increase in the percentage of errors corresponds to lower accuracy of prediction. When correlation is positive, in contrast, a growing error rate produces higher accuracy. These results indicate that correlation is an important factor that can increase the accuracy of prediction, or decrease it.

Scenario #1 reflects a situation in which the correlation between the disturbance and the measurement error is positive, while β=0.5 is positive too. Therefore, higher data accuracy can, in fact, lower the accuracy of predictions, depending on the magnitude of a respective change in covariance.

FIGURE 3.2: MAPE (Mean Absolute Percentage Error) vs. error characteristics

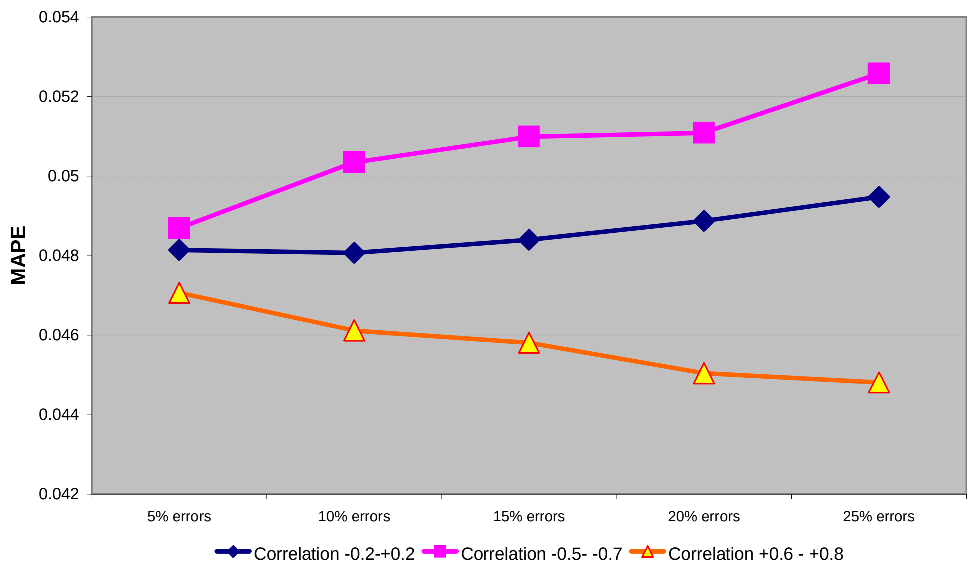
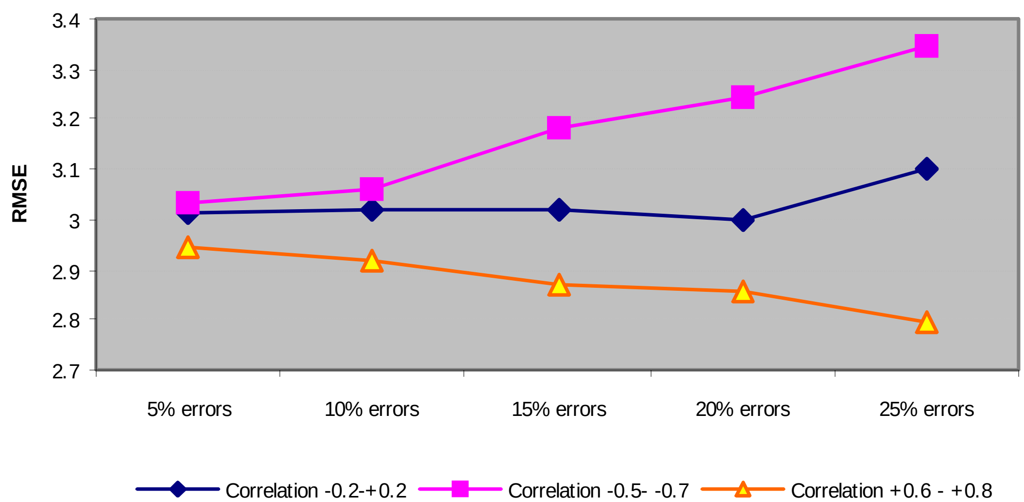


FIGURE 3.3: RMSE (Root Mean Square Error) vs. error characteristics



3.6 CONCLUSIONS

Many MIS studies reflect the belief that, setting costs aside, improvements in data accuracy are always beneficial. The findings of this inquiry challenge the validity of such belief in the more general case.

This research accounts for both stochastic and deterministic relationships between the phenomena designated by the input and output of an information system. Chapter 3 investigates the stated research questions under the condition that the information system uses a single input source for generating the output information.

The Monotonicity Theorem asserts that a monotonic increasing relationship between the accuracy of the input and the accuracy of the output is guaranteed to exist under a certain condition. This condition is satisfied when the relationship between the phenomena designated by the input and output is deterministic. More generally, however, the condition of the Monotonicity Theorem can be violated, and higher input accuracy can translate to lower output accuracy. In these cases, the study suggests, current conceptions of accuracy miss a contextual factor that can affect accuracy both ways. This factor is *dependence between errors*.

The Monotonicity Theorem can also be interpreted from an economic perspective. The condition of the Monotonicity Theorem guarantees that higher accuracy (and maximal expected utility) of the input of an information system translates to higher maximal expected utility of the output of the system. When this condition is not satisfied, higher input accuracy can translate to lower maximal expected utility.

4. ASSESSING THE INFORMATIVENESS AND VALUE OF INFORMATION INTEGRATION

4.1 INTRODUCTION

Information systems are often viewed as production systems, whose product is information. In many cases the production process combines two or more inputs via arithmetic, logical, and/or other operations—combination or integration is an essential component of the production process. The progress in communication capacity has enabled an increasingly broader scope of integration. Substantial financial investments in information integration technologies have been made, and today, data are routinely pooled from multiple systems and physical locations, and integrated in creative ways for various decision-making purposes (Stallings, 2003; Garcia-Molina et al., 2001).

The technical challenge and high cost that often characterize information system integration and distributed information systems encouraged MIS research on the economic angle of information integration. At present there exists a substantial body of MIS literature on the economic implications and value of information integration. MIS researchers studied the economic implications of information integration from an organizational and market structure perspectives, based on transaction cost theory, agency theory, organizational information processing theory, and other theories (e.g., Malone et al., 1987; Bakos, 1991; Gurbaxani and Whang, 1991; Clemons and Row, 1992; Goodhue et al., 1992). Other researchers offered analytical models of the value of information integration in supply chain settings, from an OR/MS perspective.

Unlike streams of research that make strong assumptions on the decision-making domain of information integration, e.g., supply-chain scenarios, this research develops a theory that may help decision making on information integration regardless of domain. The chosen approach links the economic value of information integration to potential improvements in information quality.

This chapter offers a formal theory of information integration. The framework in use classifies information integration situations based on two information quality characteristics—*informativeness* and *dependence*. More precisely, this study categorizes integration scenarios employing two tests, which take for granted the state set specified by the information. (1) Whether or not the input information sources can be ranked in terms of their *informativeness* for the problem at hand, and (2) whether or not the information sources are statistically *state-conditionally independent*. Four categories are created in this way.

The table below summarizes the high level findings of this research. These findings, and more, will be presented in the following sections.

TABLE 4.2: Summary of high-level results

	ISs are orthogonal	ISs are *not* orthogonal
One IS is more informative (or ISs equivalent)	-INCREASE IS NOT GUARANTEED - MONOTONICITY - PERFECT INFORMATION IS NOT ACHIEVABLE	- INCREASE IS NOT GUARANTEED - NO MONOTONICITY - PERFECT INFORMATION IS ACHIEVABLE
None is more informative	-INCREASE IS GUARANTEED - MONOTONICITY - PERFECT INFORMATION IS ACHIEVABLE	- INCREASE IS GUARANTEED - NO MONOTONICITY - PERFECT INFORMATION IS ACHIEVABLE

The analysis in this chapter is also relevant to the "monotonicity" question. Viewed from that standpoint, Chapter 4 continues the study in Chapter 3 and explores the monotonicity issue when, instead of a single source, multiple data sources are synthesized in the production of the output.

As in Chapter 3, the term information quality refers to accuracy and informativeness. The formal analysis refers mostly to the notion of "more informative", and a stronger variation on this notion, "strictly more informative", introduced in this chapter. Yet, the analysis can be adjusted to apply to the special cases of "more accurate" and "strictly more accurate". The link between accuracy / informativeness and maximal economic payoff was established in Chapter 3 through Blackwell's Thereom.

Chapter 4 is organized in the following way: Sections 4.2 and 4.3 provide a conceptual foundation for the ensuing investigation. Section 4.4 presents a basic result obtained by

Ahituv and Ronen (1988) on information integration, which is key to the rest of the paper. Sections 4.5-4.8 establish the major findings of this research—a series of theorems that are summarized by the table above. The organization of the presentation in these sections matches the organization by quadrants in Table 4.1. (Some of the theorems that apply to two quadrants are presented twice—once for each relevant quadrant; proofs are given once). Section 4.9 conducts a preliminary discussion of the implications of this study to ensemble learning algorithms research in computer science. Section 4.10 concludes the paper.

4.2 CONCEPTUAL FOUNDATIONS: INFORMATION, INTEGRATION, INFORMATIVENESS, AND ORTHOGONALITY

4.2.1 The concept of information integration

The term information integration is associated with a function that maps the values of a multi-dimensional random variable to the values of a one-dimensional random variable. Such interpretation points to the typical purpose of information integration technologies. However, it is also motivated by a major role of information systems in general: output is often produced by combining, or integrating, multiple inputs, through arithmetic, logical, and/or other operations.

A given set of sources can, nonetheless, be integrated in many ways. This study characterizes the *maximal* informativeness and economic value of information integration under various conditions. Therefore, the notion of information integration designates functions that are optimal in the sense that the informativeness and value of their output are maximal among all possible functions from the values of a given multi-dimensional random variable. Data or information are modeled, as in Chapter 3, by an information structure (IS). The formal definition of information integration is equal to the definition of a *general product* proposed by Ahituv and Ronen (1988). Definition 4.2.1.1 implies the understanding that integration corresponds to a one-to-one mapping from a given domain. Intuition directs that when information integration does not match a one-to-one function, it may produce fewer details and therefore may be less informative than an integration that matches a one to-one function. Lemma 4.2.1.1 proves this intuition.

Definition (4.2.1.1): *Integration.* $\boldsymbol{A}$ and $\boldsymbol{B}$ denote ISs defined on $S\text{x}Y_{\mathrm{A}}$ and $S\text{x}Y_{\mathrm{B}}$, respectively. $\boldsymbol{G}$ is an *integration* of $\boldsymbol{A}$ and $\boldsymbol{B}$, denoted $\boldsymbol{G}=\boldsymbol{A}@\boldsymbol{B}$, if $\boldsymbol{G}:S\text{x}Y_{\mathrm{G}}\rightarrow[0,1]$, $|Y_{\mathrm{G}}|=|Y_{\mathrm{A}}\text{x}Y_{\mathrm{B}}|$, is described by a matrix whose elements are as follows. For every $k=1,\ldots,|S|$,

$$\boldsymbol{G}_{k,1}=\Pr(y^{\mathrm{A}}_{1}\cap y^{\mathrm{B}}_{1}|s_k)$$

$$\boldsymbol{G}_{k,2}=\Pr(y^{\mathrm{A}}_{1}\cap y^{\mathrm{B}}_{2}|s_k)$$

•

•

$$\boldsymbol{G}_{k,|\mathrm{Y}_{\mathbf{B}}|}=\Pr(y^{\mathrm{A}}_{1}\cap y^{\mathrm{B}}_{|\mathrm{Y}_{\mathbf{B}}|}|s_k)$$

$$\boldsymbol{G}_{k,|\mathrm{Y}_{\mathbf{B}}+1|}=\Pr(y^{\mathrm{A}}_{2}\cap y^{\mathrm{B}}_{1}|s_k)$$

•

•

$$\boldsymbol{G}_{k,|\mathrm{Y}_{\mathbf{B}}\text{ x }\mathrm{Y}_{\mathbf{A}}|}=\Pr(y^{\mathrm{A}}_{|\mathrm{Y}_{\mathbf{A}}|}\cap y^{\mathrm{B}}_{|\mathrm{Y}_{\mathbf{B}}|}|s_k)$$

◙

G is an IS (Ahituv and Ronen, 1988). Ahituv and Ronen (1988) list properties of the integration operation, one of which is that it is associative, i.e., ***A***@(***B***@***C***)=(***A***@***B***)@***C***.

There is an infinite number of possible integrations for a given set of ISs, such that, unless additional understanding enables direct calculation (e.g., see *orthogonal ISs* below), the integration matrix is assessed based on empirical evidence. In contrast, given an integration, the input IS matrices can be derived through aggregation of columns.

Lemma 4.2.1.1: ***A***, ***B***, denote ISs defined on SxY_A and SxY_B, respectively. Let ***G*** denote their integration, ***G***=***A***@***B***, defined on SxY_G. Let ***E*** denote an IS defined on SxY_E such that:

1. $|Y_E|\leq|Y_G|$
2. There exists a partition P of Y_G and a one-to-one mapping h:$Y_E \rightarrow P$ onto P, such that a value in row i column j of ***E*** is equal to the sum of the values in row i of each of the columns of ***G*** that belong to $h(y^E_j)$

Then ***G*** is GMI than ***E***.

Proof:

By definition, ***G*** is GMI than ***E*** if there exists a Markov matrix ***M*** such that ***GM***=***E***. This proof will construct the desired ***M***.

M is defined as follows: it has $|Y_G|$ rows and $|Y_E|$ columns. $m_{i,j}$, the value in row i column j of ***M***, is equal one if $y^G_i \in h(y^E_j)$. Otherwise $m_{i,j}$=0.

This definition satisfies the requirements on $\boldsymbol{M}$. First, $\boldsymbol{M}$ is a Markov matrix. By the definition of a partition, for every $1 \le k \le |Y_G|$, $y^G{}_k$ is a member of exactly one subset, i.e., one element of P. Since h is a function onto P, and is also a one-to-one mapping, there is exactly one j such that $y^G{}_k \in h(y^E{}_j)$. Therefore the kth row of $\boldsymbol{M}$ has exactly one value which is equal one. The rest of the values are zero by definition.

Second, $\boldsymbol{GM}=\boldsymbol{E}$. $\boldsymbol{M}$ has the required dimensions—the number of its rows, $|Y_G|$, is the same as the number of columns of $\boldsymbol{G}$; the number of its columns, $|Y_E|$, is the same as the number of columns of $\boldsymbol{E}$. Furthermore, the definition of $\boldsymbol{M}$ and the definition of the matrix multiplication operation imply that the value in row i column j of $\boldsymbol{E}$ is equal to the sum of values in row i of the columns of $\boldsymbol{G}$ belonging to $h(y^E{}_j)$.

◙

Example 4.2.1.1: Consider an information system, which, similar to Section 3.2, receives data about the number of units ordered from a product and the price per-package, and computes the revenue on an order. Unlike example 3.2.1.2, price per-package can vary from one customer to another.

Input sources will be modeled first.

Suppose that the possible prices are given by S={\$0.80, \$0.95, \$1.00}, and price data suffer from recording errors. When Y={0.80, 0.95, 1.00}, an IS that describes the information that price data supply about the actual price can be the following:

(4.1)

Signal /State	**"0.80"**	**"0.95"**	**"1.00"**
$0.80	0.99	0.004	0.002
$0.95	0.003	0.99	0.003
$1.00	0.002	0.004	0.99

Suppose that there are three different packages, as in Section 3.2, so the state set associated with the revenues is *S*={$80, $95, $100, $800, $950, $1000, $8000, $9500, $1000}. If price data are used alone for informing about revenues, the respective IS, denoted next ***A***, would be the following:

(4.2)

Signal /State	**"0.80"**	**"0.95"**	**"1.00"**
$80	0.99	0.004	0.002
$95	0.003	0.99	0.003
$100	0.002	0.004	0.99
$800	0.99	0.004	0.002
$950	0.003	0.99	0.003
$1000	0.002	0.004	0.99
$8000	0.99	0.004	0.002
$9500	0.003	0.99	0.003
$10000	0.002	0.004	0.99

The values in matrices 4.1 and 4.2 reflect an assumption that, given the actual price, the signal about the price is independent of the revenue. For example, the probability of signal "0.80" is the same when revenue is $80, $800, and $8000 (0.99).

A similar line of thought suggests that an IS ***B*** that depicts the information that data on the number of units ordered supplies about the revenue can be:

(4.3)

Signal /State	“100”	“1000”	“10000”
$80	0.97	0.02	0.01
$95	0.97	0.02	0.01
$100	0.97	0.02	0.01
$800	0.01	0.98	0.01
$950	0.01	0.98	0.01
$1000	0.01	0.98	0.01
$8000	0.01	0.01	0.98
$9500	0.01	0.01	0.98
$10000	0.01	0.01	0.98

See IS 3.1.

The integration ***G*** of ***A*** and ***B*** is shown under 4.4 below. ***G*** is defined on the state set *S* as above, and the signal set *Y*={”80”, “95”, “100”, “800”, “950”, “1000”, “8000”, “9500”, “10000”}. Matrix 4.4 reveals an assumption that, given any revenue, the number of units ordered signal is conditionally independent of the price signal:

(4.4)

Signal /State	”80”	”800”	”8000”	”95”	”950”	”9500”	”100”	”1000”	”10000”
$80	.9603	.0198	.0099	.00388	.00008	.00004	.00194	.00004	.00002
$95	.00291	.00006	.00003	.9603	.0198	.0099	.00291	.00006	.00003
$100	.00194	.00004	.00002	.00388	.00008	.00004	.9603	.0198	.0099
$800	.0099	.9702	.00004	.00004	00392	.00003	.00002	.00196	.00003
$950	.00003	.00294	.00002	.0099	.9702	.00004	.00003	.00294	.0099
$1000	.00002	.00196	.0099	.00004	.00392	.00004	.0099	.9702	.00002
$8000	.0099	.0099	.9702	.00004	.00004	.00392	.00002	.00002	.00196
$9500	.00003	.00003	.00294	.0099	.0099	.9702	.00003	.00003	.00294
$10000	.00002	.00002	.00196	.00004	.00004	.00392	.0099	.0099	.9702

◘

4.2.2 Evaluating and comparing information sources

Blackwell's sufficiency criterion provides a practical way to handle the ordering of ISs. However, the ordering relation that it establishes on the set of ISs defined on a state set S is not total. This investigation examines both the condition in which the ISs that serve as inputs to an integration are ranked (also termed here "GMI-related"), and the condition in which they cannot be ranked.

Blackwell's sufficiency criterion is also weak in the sense that the superiority that it implies is not strict. Consequently, the definitions of GMI and *more accurate* allow a situation in which each IS is superior to the other, i.e., the ISs are informatively equivalent. The concept of *strictly more informative* (GSMI) is a variation on the notion of GMI that removes the described ambiguity.

Definition (4.2.2.1): *Strictly more informative.* An IS $\boldsymbol{A}$ is *strictly more informative*, or *generally and strictly more informative* (GSMI), than an IS $\boldsymbol{B}$ if $\boldsymbol{A}$ is GMI than $\boldsymbol{B}$ and $\boldsymbol{B}$ is not GMI than $\boldsymbol{A}$ (i.e., $\boldsymbol{A}$ and $\boldsymbol{B}$ are not informatively equivalent). ◙

GSMI is a transitive relation, but it is not reflexive like GMI.

When one IS is GSMI than another, it is also GMI than the other IS. Therefore, Blackwell Theorem guarantees the economic superiority of such IS.

A notion of *strictly more accurate* can be immediately defined analogous to GSMI.

Definition (4.2.2.2): *Strictly more accurate.* An IS ***A*** is *strictly more accurate* than an IS ***B*** if ***A*** is more accurate than ***B***, and ***B*** is not more accurate than ***A***. ◙

The following analysis concentrates on the concepts of GSMI and GMI rather then their special cases—more accurate and strictly more accurate.

4.2.3 State-conditional independence

"Ordinary" unconditional statistical independence requires that the probability that a random variable takes a particular value does not depend on the value of another variable. In contrast, state-conditional independence (also called *class-conditional independence*) requires that, *given any state*, the probability that a random variable takes a particular value does not depend on the value of another variable. When the probability distribution is the same regardless of state these two types of statistical independence go together. However, in the more general case they are usually contradictory (Barabash, 1965; Hussain, 1972).

Ahituv and Ronen (1988) introduce the notion of *orthogonal information structures*.

Definition (4.2.3.1): *Orthogonal information structures.* ***A*** and ***B*** denote two ISs defined on $S \times Y_A$ and $S \times Y_B$, respectively. Signals $y^A \in Y_A$, $y^B \in Y_B$, are orthogonal if $\Pr(y^A|s,y^B)=\Pr(y^A|s)$, $\forall s \in S$. ***A*** and ***B*** are orthogonal if every pair of signals $y^A \in Y_A$, $y^B \in Y_B$, is orthogonal. ◙

When two ISs are orthogonal, their integration can be directly derived from their values. The reason is that conditional joint probabilities are, in this case, equal to the product of the individual conditional probabilities. Consequently, an assumption of orthogonality is convenient in the sense that it allows easy assessment of the outcome of integration.

4.3 CONCEPTUAL FOUNDATIONS: CRITERIA USED TO CLASSIFY THE OUTCOME OF INTEGRATION

The definitions in this section establish three formal criteria on the informativenes and economic value of integration, which will be investigated in the succeeding analysis.

1. Whether or not an increase in informativeness is guaranteed through integration.
2. Whether or not higher informativeness of the individual ISs guarantees higher informativeness of their integration (monotonicity).
3. Whether or not perfect information is achievable through integration.

Each of the criteria will be defined in a formal manner. In addition, each will be briefly justified from the standpoint of its interest or usefulness for managerial decision-making.

Two special types of ISs that will be used in the formal definitions in this section and elsewhere in this study will be introduced first. These are *perfect IS*, and *null IS*.

4.3.1 Perfect IS and null IS

A perfect IS refers to an IS that offers perfect information about the state (see also *maximal information matrix*, Marschak, 1971).

Definition (4.3.1.1): *Perfect information structure*. An IS $\boldsymbol{A}$ defined on SxY_A is a *perfect IS* if it is GMI than an IS $\boldsymbol{I}$ defined on SxY_I, described by an identity matrix[2].◙

An IS $\boldsymbol{I}$ as above is GMI than any IS defined on S, i.e., $\boldsymbol{I}$ has maximal ranking. In particular, for any IS $\boldsymbol{A}$ defined on S, a Markov matrix $\boldsymbol{M}$ such that $\boldsymbol{IM}=\boldsymbol{A}$ is $\boldsymbol{M}=\boldsymbol{A}$, because $\boldsymbol{IA}=\boldsymbol{A}$. Therefore, if—as stated in Definition 4.3.1.1—$\boldsymbol{A}$ is GMI than $\boldsymbol{I}$, then $\boldsymbol{A}$ and $\boldsymbol{I}$ are informatively equivalent. Due to the transitivity of the GMI relation, $\boldsymbol{A}$ has maximal ranking too. Unlike the notion of an identity matrix, which is limited to square matrices, the notion of a perfect IS enables to study perfect information regardless of the dimensions of the IS matrix.

<u>Example 4.3.1.1:</u> Consider the IS below (4.5), denoted $\boldsymbol{A}$:

(4.5)

$$\begin{array}{r|ccc|}
\textbf{Signal /State} & \multicolumn{1}{c}{y^A_1} & y^A_2 & \multicolumn{1}{c}{y^A_3} \\
s_1 & .9 & 0 & .1 \\
s_2 & 0 & 1 & 0
\end{array}$$

Intuition says that $\boldsymbol{A}$ provides perfect information, since each of the possible signals points, with certainty, to one state. For example, when y^A_1 is observed, it implies s_1 with certainty. In agreement with this intuition, $\boldsymbol{A}$ is GMI than an IS $\boldsymbol{I}$ whose matrix is the identity matrix. A Markov matrix $\boldsymbol{M}$ such that $\boldsymbol{AM}=\boldsymbol{I}$ is the following:

---

[2] An identity matrix is a square matrix whose diagonal elements are 1s and whose off-diagonal elements are 0s.

(4.6)

Signal /Signal	y^{I}_1	y^{I}_2
y^{A}_1	1	0
y^{A}_2	0	1
y^{A}_3	1	0

Therefore, Definition 4.3.1.1 says that $\boldsymbol{A}$ is a perfect IS.

◘

Lemma 4.3.1.1: An IS $\boldsymbol{A}$ is a perfect IS if and only if it has at most one positive value in each column of the matrix.

Proof:

The proof, by contradition, assumes that an IS has at least one positive value in each column. If this is not true, then, since a Markov matrix has only non-negative values, one or more columns have only zero values. Such column(s) can be eliminated and the signal set can be reduced accordingly.

(⇒)

Suppose that an IS $\boldsymbol{A}$ defined on S is a perfect IS, and that, contrary to the theorem, $\boldsymbol{A}$ has more than one positive value in one or more of its columns. Especially, it will be assumed—without loss of generality—that the lth column of $\boldsymbol{A}$ has two positive values, in rows k and j.

By definition, since ***A*** is a perfect IS, there exists a Markov matrix ***M*** such that ***AM***=***I***, ***I*** is an IS defined on *S* and expressed by an identity matrix. Since ***M*** is a Markov matrix, its values are always non-negative, and at least one value in each row of ***M*** is positive. It can be assumed—without loss of generality—that m_{li}, the value in the *l*th row, column *i*, of ***M***, is positive. ***A*** is a Markov matrix too, so the values of ***A*** are all non-negative. Therefore, ***I*** must be such that the value in column *i*, row *k*, of ***I***, which is derived from row *k* of ***A*** and column *i* of ***M***, is positive. Similarly, the value in column *i*, row *j*, of ***I***, which is derived from row *j* of ***A*** and column *i* of ***M***, is positive. It follows that the *i*th column of ***I*** has two positive values. But this contradicts the assumption that ***I*** is an identity matrix. Therefore, ***A*** has exactly one positive value in each column of the matrix.

($\Leftarrow$)

Suppose that an IS ***A*** has a special structure, i.e., exactly one positive value in each column of the matrix. It will be shown that there exists a Markov matrix ***M*** such that ***AM***=***I*** . Therefore, by definition, ***A*** is a perfect IS.

M will be defined as follows: The number of columns in ***M*** is the same as the number of rows in ***A***, i.e., |*S*|. The number of rows is the same as the number of columns in ***A***. For every row *j* of ***A***, column *j* of ***M*** is defined such that its *i*th element is equal one if there is a positive value in row *j*, column *i*, of ***A***, and zero otherwise.

M is Markov matrix. Since, by assumption, each column of ***A*** has exactly one positive value, each row of ***M*** has exactly one positive value; such value is equal one by the definition of ***M***. The other values in the row are zero.

$\boldsymbol{AM}=\boldsymbol{I}$. By its definition, $\boldsymbol{M}$ has the required dimensions for such product. The number of rows in $\boldsymbol{M}$ is the same as the number of columns in $\boldsymbol{A}$. The number of columns is the same as the number of columns in $\boldsymbol{I}$. In addition, the sum of the values in any row of $\boldsymbol{A}$ is equal one since $\boldsymbol{A}$ is a Markov matrix, and, by the definition of $\boldsymbol{M}$, the ith element of column j in $\boldsymbol{M}$ has the value one if there is a positive value in the ith element of row j in $\boldsymbol{A}$. Therefore, by the definition of the matrix multiplication operation, a product of row j of $\boldsymbol{A}$ and column j of $\boldsymbol{M}$ amounts to summing up the values in row j of $\boldsymbol{A}$, i.e., it will yield one. A product of row j of $\boldsymbol{A}$ and column $k \neq j$ of $\boldsymbol{M}$ will yield zero. The reason is that if row j of $\boldsymbol{A}$ has a positive value in column l, then column j of $\boldsymbol{M}$ has, by definition, value 1 in its lth row, and since $\boldsymbol{M}$ is a Markov matrix, row l has value zero in any column $k \neq j$. Therefore, when row j of $\boldsymbol{A}$ is multiplied by any column $k \neq j$ of $\boldsymbol{M}$, the positive value in column l of this row is multiplied by zero.

◙

Another special IS is a *null IS*. A null IS refers to an IS that offers no information at all about the state (see also *minimal information matrix*, Marschak, 1971).

Definition (4.3.1.2): *Null information structure*. An IS $\boldsymbol{A}$ is a *null IS* if the rank of the matrix is 1.◙

When an IS $\boldsymbol{A}$ meets the condition of Definition 4.3.1.2, it can be shown that any IS defined on S is GMI than $\boldsymbol{A}$, i.e., $\boldsymbol{A}$ has minimal ranking.

Example 4.3.1.2:

(4.7)

Signal /State	y^A_1	y^A_2
s_1	.9	.1
s_2	.9	.1

IS 4.7 is a null IS, since its columns (/rows) are linearly dependent, i.e., the rank of the matrix is equal one. Intuition says that IS 4.7 offers no information at all. When either y^A_1 or y^A_2 is observed, the IS shows total uncertainty as to which of s_1 and s_2 occurs—the signal can be equally issued given any of the states.

◘

4.3.2 Increase in quality is guaranteed

The definition below formalizes the situation in which the informativeness of the integration is strictly higher than the informativeness of any of the input ISs.

Definition (4.3.2.1): *Increase in informativeness is guaranteed.* ***A*** and ***B*** denote two ISs defined on SxY_A and SxY_B, respectively. An *increase in informativeness is guaranteed* for ***A*** and ***B*** if their integration ***G***=***A***@***B*** is GSMI than ***A*** and ***B***.◙

Definition (4.3.2.2): *Increase in informativeness is not guaranteed.* Let I denote a set of integrations defined on state set S. If there exists an integration ***G***=***A***@***B*** in I such that $\boldsymbol{G}$ is not GSMI than $\boldsymbol{A}$ and $\boldsymbol{B}$, then an *increase in informativeness is not guaranteed* in I. ◙

As mentioned earlier, the input IS matrices can be uniquely derived from their integration, such that when $\boldsymbol{G}$ is given, it is assumed that $\boldsymbol{A}$ and $\boldsymbol{B}$ are given too.

If a set of integrations is such that some do not entail strictly higher informativeness, then aiming at this set is risky—the return may be zero. Everything else equal, a decision that avoids this set may produce higher value.

4.3.3 Monotonicity

Monotonicity is defined as follows:

Definition (4.3.3.1): *Monotonicity*. $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$, $\boldsymbol{D}$, are ISs defined on SxY_A, SxY_B, SxY_C, and SxY_D, respectively. Let $\boldsymbol{G}^{AB}$ denote the integration of $\boldsymbol{A}$ and $\boldsymbol{B}$ ($\boldsymbol{G}^{AB}$=$\boldsymbol{A}$@$\boldsymbol{B}$). Let $\boldsymbol{G}^{CD}$ denote the integration of $\boldsymbol{C}$ and $\boldsymbol{D}$ ($\boldsymbol{G}^{AC}$=$\boldsymbol{A}$@$\boldsymbol{C}$). If $\boldsymbol{C}$ is GMI than $\boldsymbol{A}$ and $\boldsymbol{D}$ is GMI than $\boldsymbol{B}$ implies $\boldsymbol{G}^{CD}$ is GMI than $\boldsymbol{G}^{AB}$, then the relationship between the informativeness of of $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$, $\boldsymbol{D}$, and the informativeness of $\boldsymbol{G}^{CD}$ and $\boldsymbol{G}^{AB}$ is *monotonic*.◙

Definition (4.3.3.2): *Non-monotonicity*. Let I denote a set of integrations defined on a state set S. If there exist integrations $\boldsymbol{G}^{AB}$=$\boldsymbol{A}$@$\boldsymbol{B}$, $\boldsymbol{G}^{CD}$=$\boldsymbol{C}$@$\boldsymbol{D}$, in I, such that $\boldsymbol{C}$ is GMI than $\boldsymbol{A}$, $\boldsymbol{D}$ is GMI than $\boldsymbol{B}$, and $\boldsymbol{G}^{CD}$ is not GMI than $\boldsymbol{G}^{AB}$, then the relationship between the informativeness of the ISs and the informativeness of their integration is *not monotonic* in I.◙

Monotonicity is a convenient property—the decision-maker is assured that higher informativeness of individual sources will enhance their integration. This way, for example, monotonicity can serve to justify investments in higher data quality where integration is an issue. The knowledge that higher quality of individual sources has positive link to their integration also simplifies the identification of "good" sources in other integration situations. It suggests a simple source selection strategy: select the sources with the highest individual informativeness. In fact, assessments that are based on an implicit assumption of monotonicity are common.

However, if monotonicity does not hold, or may not hold, decision processes should better take this condition into account. Otherwise, searches for highly valuable integrations can miss their target if they follow the strategy of selecting the sources with the highest informativeness. Similarly, investments in data quality improvement may turn to be less valuable than estimated because of negative effects on associated integrations.

4.3.4 Perfect information is achievable

The notion of *perfect information is achievable* points to any set of integrations defined on a common state set, in which some or all the members are perfect ISs. A decision-making process may benefit from aiming at such integration set.

Definition (4.3.4.1): *Perfect information is not achievable*. $\boldsymbol{A}$ and $\boldsymbol{B}$ are ISs defined on SxY_A, SxY_B, respectively. *Perfect information is not achievable* by $\boldsymbol{A}$ and $\boldsymbol{B}$ if none of $\boldsymbol{A}$, $\boldsymbol{B}$, is a perfect IS implies that $\boldsymbol{G}=\boldsymbol{A}@\boldsymbol{B}$ is not a perfect IS.◙

Definition (4.3.4.2): *Perfect information is achievable*. I is a set of integrations defined on a state set S. *Perfect information is achievable* in I if there exists ***G***=***A***@***B*** in I such that none of ***A***, ***B***, is a perfect IS and ***G***=***A***@***B*** is a perfect IS.◙

4.4 BASIC RESULT: INTEGRATION NEVER LOWERS THE QUALITY OF INFORMATION

Ahituv and Ronen (1988) prove a basic, intuitive result using the IS model: The informativeness of an integration is never lower than the informativeness of any of the input ISs.

Theorem 4.4.1. For any two ISs ***A***, ***B***, defined on $S\text{x}Y_A$, $S\text{x}Y_B$, respectively, the integration of ***A*** and ***B***, ***G***=***A***@***B***, is GMI than ***A*** and ***B***.◙

Despite the importance of this result, it marks only the beginning of an understanding. Various research directions, some explored here, may contribute new understanding on the quality and value of information integration. Particular avenues for further study are:

- It would be useful to know if informativeness always strictly increases when ISs are integrated, or whether it is possible that informativeness will not increase. Given the latter scenario, a capacity to distinguish conditions under which integration does not involve any gain can be useful.

- An ability to characterize conditions in which integration can produce perfect information can be valuable.

- The cost of implementing a theory as indicated above may be heavy. Therefore, a theory that would point to a “safe gain” strategy to integration—a simple, inexpensive approach, that will guarantee one or another measure of increase in information quality—might be appreciated.

4.5 INTEGRATION WHEN SOURCES ARE STATE-CONDITIONALLY INDEPENDENT, AND CAN BE ORDERED BY THEIR QUALITY

4.5.1 Increase in quality is not guaranteed

It will be proved that integrations of orthogonal ISs are strictly more informative than any of the input ISs, i.e., increase in informativness is guaranteed, except for special instances. Theorem 4.5.1.1 applies to orthogonal ISs in general, except for integrations of ISs such that there does not exist any pair of states for which the respective sub-matrices of both ISs are neither null ISs, nor perfect ISs. When one IS is GMI than the other, the latter situation implies that for every pair of states, the matching sub-matrix of the more informative IS describes a perfect IS, and/or, the sub-matrix of the less informative IS describes a null IS. An example suggests that these integrations are valueless. Theorem 4.7.1.1 in Section 4.7.1 provides a complete characterization of the case in which integrations are valueless.

The superior informativeness that Theorem 4.5.1.1 establishes is stronger than the one implied by Theorem 4.4.1. While Theorem 4.4.1 does not provide any guarantee that integration yields positive gain in informativeness, Theorem 4.5.1.1 does.

Theorem 4.5.1.1. ***A***, ***B***, denote ISs defined on SxY_A, SxY_B, respectively, such that ***A*** and ***B*** are orthogonal. An increase in informativeness is guaranteed for ***A*** and ***B*** if:

1. There exists a two row sub-matrix of ***B***, ***B'***, in which the rows correspond to $S' \subseteq S$, such that ***B'*** is neither a null IS nor a perfect IS
1. ***A'***, a two row sub-matrix of ***A*** whose rows correspond to $S' \subseteq S$, is neither a null IS nor a perfect IS.

Proof:

The first part of the proof is offered in order to enhance the understandability of the more general cases. It assumes that ***A*** and ***B*** are 2x2 matrices. The proof of the general case appears in the Appendix.

It will shown, by contradiction, that, under the assumed conditions, neither ***A*** nor ***B*** is GMI than ***G***=***A***@***B***, so ***G*** is GSMI than ***A*** and ***B*** by definition. Therefore, by definition, an increase in informativeness is guaranteed for ***A*** and ***B***.

Without loss of generality, suppose that, contrary to the theorem's conclusion, ***A*** is GMI than ***G***, such that there exists a Markov matrix ***M***, ***AM***=***G***.

Let

$$\boldsymbol{A}=\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \qquad \boldsymbol{B}=\begin{vmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{vmatrix}$$

Consider the integration ***G***=***A***@***B***. Based on the definitions of integration and orthogonal ISs, ***G*** is described by:

$$\boldsymbol{G}=\begin{vmatrix} a_{11}b_{11} & a_{11}b_{12} & a_{12}b_{11} & a_{12}b_{12} \\ a_{21}b_{21} & a_{21}b_{22} & a_{22}b_{21} & a_{22}b_{22} \end{vmatrix}$$

Since det $\boldsymbol{A}\neq 0$ by assumption (or else $\boldsymbol{A}$ is a null IS), a single solution to any equation set whose coefficient matrix is $\boldsymbol{A}$ is guaranteed to exist. Solving for $\boldsymbol{M}$ such that $\boldsymbol{AM}=\boldsymbol{G}$ corresponds to four equation sets, one equation set for each column of $\boldsymbol{G}$. Therefore, the existence of a single solution to any one of the four sets is guaranteed. A global analysis of the solutions will be carried next.

First, based on Cramer's rule (Simon and Blume, 1994) the solutions correspond to the following $\boldsymbol{M}$ (a 2 x 4 matrix):

$$m_{11}=\det\begin{vmatrix} a_{11}b_{11} & a_{12} \\ a_{21}b_{21} & a_{22} \end{vmatrix}\div\det \boldsymbol{A}$$

$$m_{12}=\det\begin{vmatrix} a_{11}b_{12} & a_{12} \\ a_{21}b_{22} & a_{22} \end{vmatrix}\div\det \boldsymbol{A}$$

$$m_{13}=\det\begin{vmatrix} a_{12}b_{11} & a_{12} \\ a_{22}b_{21} & a_{22} \end{vmatrix}\div\det \boldsymbol{A}$$

$$m_{14}=\det\begin{vmatrix} a_{12}b_{12} & a_{12} \\ a_{22}b_{22} & a_{22} \end{vmatrix}\div\det \boldsymbol{A}$$

$$m_{21}=\det\begin{vmatrix} a_{11} & a_{11}b_{11} \\ a_{21} & a_{21}b_{21} \end{vmatrix} \div \det \boldsymbol{A}$$

$$m_{22}=\det\begin{vmatrix} a_{11} & a_{11}b_{12} \\ a_{21} & a_{21}b_{22} \end{vmatrix} \div \det \boldsymbol{A}$$

$$m_{23}=\det\begin{vmatrix} a_{11} & a_{12}b_{11} \\ a_{21} & a_{22}b_{21} \end{vmatrix} \div \det \boldsymbol{A}$$

$$m_{24}=\det\begin{vmatrix} a_{11} & a_{12}b_{12} \\ a_{21} & a_{22}b_{22} \end{vmatrix} \div \det \boldsymbol{A}$$

Suppose that the first two elements of the first row of ***M***, i.e., $m_{1,1}$ and $m_{1,2}$, are aggregated.

B is a Markov matrix, therefore:

$m_{1,1} + m_{1,2}= [a_{1,1}a_{2,2}b_{1,1} - a_{1,2}a_{2,1}b_{2,1} + a_{1,1}a_{2,2}b_{1,2} - a_{1,2}a_{2,1}b_{2,2}]\div\det \boldsymbol{A}=[a_{1,1}a_{2,2}(b_{1,1}+b_{1,2}) - a_{1,2}a_{2,1}(b_{2,1}+ b_{2,2})] \div\det \boldsymbol{A}=[a_{1,1}a_{2,2} - a_{1,2}a_{2,1}] \div\det \boldsymbol{A}= \det \boldsymbol{A} \div\det \boldsymbol{A}=1$

Suppose next that the first two elements of the second row of ***M***, $m_{2,1}$ and $m_{2,2}$, are aggregated as well:

$m_{2,1} + m_{2,2} = [a_{1,1}a_{2,1}b_{2,1} - a_{1,1}a_{2,1}b_{1,1} + a_{1,1}a_{2,1}b_{2,2} - a_{1,1}a_{2,1}b_{1,2}] \div \det \boldsymbol{A} = [a_{1,1}a_{2,1}(b_{2,1}+b_{2,2}) - a_{1,1}a_{2,1}(b_{1,1}+ b_{1,2})] \div \det \boldsymbol{A} = [a_{1,1}a_{2,1} - a_{1,1}a_{2,1}] \div \det \boldsymbol{A} = 0$

M is a Markov matrix, therefore the elements of ***M*** are all non-negative. Therefore:

$m_{2,1}= [a_{1,1}a_{2,1}b_{2,1} - a_{1,1}a_{2,1}b_{1,1}] \div \det \boldsymbol{A} = 0$

$m_{2,2} = [a_{1,1}a_{2,1}b_{2,2} - a_{1,1}a_{2,1}b_{1,2}] \div \det \boldsymbol{A} = 0$

By assumption, $\det \boldsymbol{A}\neq 0$. Consequently:

$a_{1,1}a_{2,1}b_{2,1}$ - $a_{1,1}a_{2,1}b_{1,1} = a_{1,1}a_{2,1}(b_{2,1} - b_{1,1}) = 0$

$a_{1,1}a_{2,1}b_{2,2}$ - $a_{1,1}a_{2,1}b_{1,2} = a_{1,1}a_{2,1}(b_{2,2} - b_{1,2}) = 0$

Since det $\boldsymbol{B}\neq 0$ by assumption (or else $\boldsymbol{B}$ is a null IS), $a_{1,1}a_{2,1}=0$. This implies that either $a_{2,1}=0$ (if det $\boldsymbol{A}>0$, but then $a_{1,1}>0$), or $a_{1,1}=0$ (if det $\boldsymbol{A}<0$, but then $a_{2,1}>0$).

The last two elements of the first row of $\boldsymbol{M}$, $m_{1,3}$ and $m_{1,4}$, are aggregated too:

$m_{1,3} + m_{1,4} = [a_{1,2}a_{2,2}b_{1,1} - a_{1,2}a_{2,2}b_{2,1} + a_{1,2}a_{2,2}b_{1,2} - a_{1,2}a_{2,2}b_{2,2}] \div \det \boldsymbol{A} = [a_{1,2}a_{2,2}(b_{1,1}+b_{1,2})$ - $a_{1,2}a_{2,2}(b_{2,1}+ b_{2,2})] \div \det \boldsymbol{A} = [a_{1,2}a_{2,2} - a_{1,2}a_{2,2}] \div \det \boldsymbol{A} = 0$

Again, $\boldsymbol{M}$ is a Markov matrix, therefore:

$m_{1,3} = [a_{1,2}a_{2,2}b_{1,1} - a_{1,2}a_{2,2}b_{2,1}] \div \det \boldsymbol{A} = 0$

$m_{1,4} = [a_{1,2}a_{2,2}b_{1,2} - a_{1,2}a_{2,2}b_{2,2}] \div \det \boldsymbol{A} = 0$

det $\boldsymbol{A}\neq 0$, therefore:

$a_{1,2}a_{2,2}b_{1,1} - a_{1,2}a_{2,2}b_{2,1} = a_{1,2}a_{2,2}(b_{1,1} - b_{2,1}) = 0$

$a_{1,2}a_{2,2}b_{1,2} - a_{1,2}a_{2,2}b_{2,2} = a_{1,2}a_{2,2}(b_{1,2} - b_{2,2}) = 0$

Since det $\boldsymbol{B}\neq 0$, $a_{1,2}a_{2,2}= a_{1,2}a_{2,2}=0$, which implies that either $a_{1,2}=0$ (if det $\boldsymbol{A}>0$, but then $a_{2,2}>0$) or $a_{2,2}=0$ (if det $\boldsymbol{A}<0$, but then $a_{1,2}>0$).

Therefore, either det $\boldsymbol{A}>0$ and $a_{1,2}=0$, $a_{2,1}=0$, which is ruled out because it implies that $\boldsymbol{A}$ is a perfect IS in contradiction to the theorem's assumptions, or det $\boldsymbol{A}<0$, $a_{2,2}=0$, $a_{1,1}=0$, which is ruled out too out of the same reason. Therefore, $\boldsymbol{A}$ is not GMI than $\boldsymbol{G}$.

Since $\boldsymbol{A}$ and $\boldsymbol{B}$ are symmetric, the same proof applies to $\boldsymbol{B}$ such that $\boldsymbol{B}$ is not GMI than $\boldsymbol{G}$ either. ◙

<u>Example 4.5.1.1:</u> This example will demonstrate a situation in which two ISs are orthogonal and one is more informative than the other—consistent with the assumptions of the analysis in this section—and the ISs also satisfy the other conditions of Theorem 4.5.1.1. As predicted by the theorem, their integration has strictly higher informativness than both.

An IS ***A*** is the following:

(4.8)

Signal /State	y^A_1	y^A_2
s_1	.9	.1
s_2	.6	.4

An IS ***B*** is the following:

(4.9)

Signal /State	y^B_1	y^B_2
s_1	.57	.43
s_2	.48	.52

A is GMI than ***B***. Especially, a Markov Matrix ***M*** such that ***AM***=***B*** is:

(4.10)

Signal /Signal	y^B_1	y^B_2
y^A_1	.6	.4
y^A_2	.3	.7

None of ***A***, ***B***, is a perfect IS or a null IS. Furthermore, ***A*** and ***B*** are orthogonal. Due to their orthogonality, G^{AB}=***A***@***B*** is directly derived from ***A*** and ***B***:

(4.11)

Signal /State	y^{AB}_1	y^{AB}_2	y^{AB}_3	y^{AB}_4
s_1	0.513	0.387	0.057	0.043
s_2	0.288	0.312	0.192	0.208

It can be shown that G^{AB} is GSMI than ***A*** and ***B***, such that an increase in informativeness is guaranteed for ***A*** and ***B***. First, G^{AB} is GMI than ***A***—a Markov matrix ***M*** such that $G^{AB}M$=***A*** is:

(4.12)

Signal /Signal	y^{A}_1	y^{A}_2
Y^{AB}_1	1	0
y^{AB}_2	1	0
y^{AB}_3	0	1
y^{AB}_4	0	1

G^{AB} is also GMI than ***B***—a Markov matrix ***M*** such that $G^{AB}M$=***B*** is:

(4.13)

Signal /Signal	y^{B}_1	y^{B}_2
y^{AB}_1	1	0
y^{AB}_2	0	1
y^{AB}_3	1	0
y^{AB}_4	0	1

Second, $\boldsymbol{A}$ is not GMI than $\boldsymbol{G}^{AB}$ and $\boldsymbol{B}$ is not GMI than $\boldsymbol{G}^{AB}$. This statement can be proved by contradiction, through attempts to solve a set of equations for the values of a Markov matrix $\boldsymbol{M}$ such that $\boldsymbol{AM}=\boldsymbol{G}^{AB}$ (when proving that $\boldsymbol{A}$ is not GMI than $\boldsymbol{G}^{AB}$) or $\boldsymbol{BM}=\boldsymbol{G}^{AB}$ (when proving that $\boldsymbol{B}$ is not GMI than $\boldsymbol{G}^{AB}$).

Here is an outline of the proof: according to the dimensions of $\boldsymbol{A}$ (or $\boldsymbol{B}$) and $\boldsymbol{G}^{AB}$, $\boldsymbol{M}$ must have two rows and four columns. Therefore, the number of unknowns is eight. The equation set has ten equations. Two equations relate to the constraint on a Markov matrix that values in each row sum to one. The coefficients in the rest of the equations are derived from $\boldsymbol{A}$ in the case of $\boldsymbol{AM}=\boldsymbol{G}^{AB}$, or $\boldsymbol{B}$ in the case of $\boldsymbol{BM}=\boldsymbol{G}^{AB}$. Finally, there are also non-negativity constraints on the unknowns, i.e., eight non-negativity constraints.

It can be shown that the rank of the coefficient matrix is not the same as the rank of the augmented matrix (an augmented matrix specifies both coefficients and right-hand-side values). This implies that the equation set does not have a solution (Simon and Blume, 1994).

◘

<u>Example 4.5.1.2:</u> This example demonstrates, again, a situation in which—as assumed by the analysis in this section—two ISs are orthogonal and one has higher informativeness than the other. However, this integration agrees with the special pattern that Theorem 4.5.1.1 excludes. The example provides evidence that when, in addition to orthogonality, one IS is GMI than the other, such integration does not offer strictly higher informativeness.

An IS ***A*** is the following:

(4.14)

Signal /State	y^A_1	y^A_2	y^A_3	y^A_4
s_1	.5	0	0	.5
s_2	0	.8	.2	0
s_3	0	.8	.2	0

An IS ***B*** is the following:

(4.15)

Signal /State	y^B_1	y^B_2	y^B_3
s_1	0.9	0.1	0
s_2	0.6	0.3	0.1
s_3	0.6	0.3	0.1

A is GMI than ***B***. A Markov Matrix ***M*** such that ***AM***=***B*** is:

(4.16)

Signal /Signal	y^B_1	y^B_2	y^B_3
y^A_1	0.9	0.1	0
y^A_2	0.6	0.3	0.1
y^A_3	0.6	0.3	0.1
y^A_4	0.9	0.1	0

The ISs exhibit the properties that Theorem 4.5.1.1 rules out. The first two rows of ***A*** form a perfect IS defined on $\{s_1, s_2\}$, and the first and the third row form a perfect IS defined on $\{s_1, s_3\}$. The second and third rows of ***A*** form a null IS defined on $\{s_2, s_3\}$, and the respective sub-matrix of ***B*** is a null IS too. An intuitive explanation might say that ***A*** represents an information source that has acquired expertise within some state subsets, but has no understanding at all beyond this domain.

$\boldsymbol{A}$ and $\boldsymbol{B}$ are orthogonal, therefore, $\boldsymbol{G}^{AB}=\boldsymbol{A}@\boldsymbol{B}$ is:

(4.17)

	y^{AB}_1	y^{AB}_2	y^{AB}_3	y^{AB}_4	y^{AB}_5	y^{AB}_6	y^{AB}_7	y^{AB}_8	y^{AB}_9	y^{AB}_{10}	y^{AB}_{11}	y^{AB}_{12}
s_1	.45	.05	0	0	0	0	0	0	0	.45	.05	0
s_2	0	0	0	.48	.24	.08	.12	.06	.02	0	0	0
s_3	0	0	0	.48	.24	.08	.12	.06	.02	0	0	0

$\boldsymbol{A}$ is GMI than $\boldsymbol{G}^{AB}$, so $\boldsymbol{G}^{AB}$ does not offer any increase in informativeness compared to $\boldsymbol{A}$. In particular, a Markov matrix $\boldsymbol{M}$ such that $\boldsymbol{AM}=\boldsymbol{G}^{AB}$ is:

(4.18)

	y^{AB}_1	y^{AB}_2	y^{AB}_3	y^{AB}_4	y^{AB}_5	y^{AB}_6	y^{AB}_7	y^{AB}_8	y^{AB}_9	y^{AB}_{10}	y^{AB}_{11}	y^{AB}_{12}
y^{A}_1	.9	.1	0	0	0	0	0	0	0	0	0	0
y^{A}_2	0	0	0	.6	.3	.1	0	0	0	0	0	0
y^{A}_3	0	0	0	0	0	0	.6	.3	.1	0	0	0
y^{A}_4	0	0	0	0	0	0	0	0	0	.9	.1	0

◘

4.5.2 Monotonicity

Theorem 4.5.2.1 asserts that in integrations of orthogonal ISs, the relationship between the informativeness of the ISs and the informativeness of their integration is monotonic.

Note that $\boldsymbol{A}$ and $\boldsymbol{A'}$ in Lemma 4.5.2.1 may represent different information sources.

Lemma 4.5.2.1: $\boldsymbol{A}$, $\boldsymbol{A'}$, $\boldsymbol{B}$, $\boldsymbol{C}$ are ISs defined on $S \mathrm{x} Y_A$, $S \mathrm{x} Y_{A'}$, $S \mathrm{x} Y_B$, and $S \mathrm{x} Y_C$, respectively. $\boldsymbol{A}$ and $\boldsymbol{A'}$ have equal matrices, $\boldsymbol{A}$ and $\boldsymbol{B}$ are orthogonal, and $\boldsymbol{A'}$ and $\boldsymbol{C}$ are orthogonal. Let $\boldsymbol{G}^{AB}$ denote the integration of $\boldsymbol{A}$ and $\boldsymbol{B}$ ($\boldsymbol{G}^{AB}=\boldsymbol{A}@\boldsymbol{B}$). Let $\boldsymbol{G}^{A'C}$ denote the integration of $\boldsymbol{A'}$ and $\boldsymbol{C}$ ($\boldsymbol{G}^{A'C}=\boldsymbol{A'}@\boldsymbol{C}$). Then $\boldsymbol{C}$ is GMI than $\boldsymbol{B}$ implies $\boldsymbol{G}^{A'C}$ is GMI than $\boldsymbol{G}^{AB}$.

Proof:

Suppose that $\boldsymbol{C}$ is GMI than $\boldsymbol{B}$. It will be shown, by construction, that there is a Markov matrix $\boldsymbol{M}$ such that $\boldsymbol{G}^{A'C}\boldsymbol{M}=\boldsymbol{G}^{AB}$, therefore $\boldsymbol{G}^{A'C}$ is GMI than $\boldsymbol{G}^{AB}$.

Since $\boldsymbol{C}$ is GMI than $\boldsymbol{B}$, then, by definition, there is a Markov matrix $\boldsymbol{M^*}$ such that $\boldsymbol{CM^*}=\boldsymbol{B}$. $\boldsymbol{M}$ will be defined based on $\boldsymbol{M^*}$ in the following manner. $\boldsymbol{M}$ is block diagonal; each block along the diagonal is a copy of $\boldsymbol{M^*}$, and the other elements are zero; the number of blocks is the same as the number of columns in $\boldsymbol{A'}$, i.e., $|Y_{A'}|$

$\boldsymbol{M}$ is a Markov matrix, because each of its rows consists of a row of $\boldsymbol{M^*}$, which itself is a Markov matrix, "wrapped" with zeros. In addition, $\boldsymbol{M}$ has the dimensions as required for $\boldsymbol{G}^{A'C}\boldsymbol{M}=\boldsymbol{G}^{AB}$. The number of rows in $\boldsymbol{M}$ is equal the number of columns in $\boldsymbol{G}^{A'C}$, $|Y_C|\cdot|Y_{A'}|$, since the number of rows in a block, i.e., in $\boldsymbol{M^*}$, is $|Y_C|$ and the number of blocks in $\boldsymbol{M}$ is $|Y_{A'}|$. The number of columns in $\boldsymbol{M}$ is equal the number of columns in $\boldsymbol{G}^{AB}$, $|Y_B|\cdot|Y_A|$, since the number of columns in a block i.e., in $\boldsymbol{M^*}$, is $|Y_B|$, the number of blocks is $|Y_{A'}|$, and $|Y_{A'}|=|Y_A|$ by an assumption of the theorem.

It can be further shown that $\boldsymbol{G}^{A'C}\boldsymbol{M}=\boldsymbol{G}^{AB}$. In particular, it will be shown that the product of the first row of $\boldsymbol{G}^{A'C}$ and the first column of $\boldsymbol{M}$ is equal to the value in the first column and row of $\boldsymbol{G}^{AB}$. The same logic can be applied for any row of $\boldsymbol{G}^{A'C}$ and column of $\boldsymbol{M}$.

Let $y^{A}_{1}, y^{A}_{2},..,y^{A}_{|Y_A|}$ denote the elements of Y_A. Let $y^{A'}_{1}, y^{A'}_{2},..,y^{A'}_{|Y_{A'}|}$ denote the elements of $Y_{A'}$. Similarly, $y^{B}_{1}, y^{B}_{2},..,y^{B}_{|Y_B|}$ will denote the elements of Y_B; $y^{C}_{1}, y^{C}_{2},..\ y^{C}_{|Y_C|}$ will denote the elements of Y_C. Also, let m_{ij} denote the value in row i, column j, of $\boldsymbol{M}$.

Consider the first row of $\boldsymbol{G}^{A'C}$. Due to the assumption of orthogonality, the elements of this row have the form $\Pr(y^{A'}_{1}|s)\cdot\Pr(y^{C}_{1}|s)$, $\Pr(y^{A'}_{1}|s)\cdot\Pr(y^{C}_{2}|s)$,......,$\Pr(y^{A'}_{2}|s)\cdot\Pr(y^{C}_{1}|s)$, $\Pr(y^{A'}_{2}|s)\cdot\Pr(y^{C}_{2}|s)$,….,$\Pr(y^{A'}_{|Y_A|}|s)\cdot\Pr(y^{C}_{|Y_C|}|s)$.

By assumption, however, the matrices of $\boldsymbol{A}$, $\boldsymbol{A'}$, are equal. Therefore, the above can be re-written as:

$\Pr(y^{A}_{1}|s)\cdot\Pr(y^{C}_{1}|s)$,$\Pr(y^{A}_{1}|s)\cdot\Pr(y^{C}_{2}|s)$,..,$\Pr(y^{A}_{2}|s)\cdot\Pr(y^{C}_{1}|s)$,$\Pr(y^{A}_{2}|s)\cdot\Pr(y^{C}_{2}|s)$,….,$\Pr(y^{A}_{|Y_A|}|s)\cdot \Pr(y^{C}_{|Y_C|}|s)$.

A product of the first row of $\boldsymbol{G}^{A'C}$ and the first column of $\boldsymbol{M}$ will therefore have the form:

$\Pr(y^{A}_{1}|s)\cdot\Pr(y^{C}_{1}|s)\cdot m_{11}+\Pr(y^{A}_{1}|s)\cdot\Pr(y^{C}_{2}|s)\cdot m_{21}+\ldots+\Pr(y^{A}_{2}|s)\cdot\Pr(y^{C}_{1}|s)\cdot m_{|Y_C|+1,1}+\ldots \quad \ldots+ \Pr(y^{A}_{|Y_A|}|s)\cdot \Pr(y^{C}_{|Y_C|}|s)\cdot m_{|Y_C|\cdot|Y_A|,1}$

In fact, however, since $\boldsymbol{M}$ is block diagonal, all the values in its first column are zero except perhaps to the first $|Y_C|$ elements. Therefore the above product can be re-written as:

$\Pr(y^A{}_1|s)\cdot\Pr(y^C{}_1|s)\cdot m_{11} + \Pr(y^A{}_1|s)\cdot\Pr(y^C{}_2|s)\cdot m_{21} + \Pr(y^A{}_1|s)\cdot\Pr(y^C{}_3|s)\cdot m_{31}+\ldots + \Pr(y^A{}_1|s)\cdot \Pr(y^C{}_{|Y_C|}|s)\cdot m_{|Y_C|,1}=$

$\Pr(y^A{}_1|s)\cdot\{\Pr(y^C{}_1|s)\cdot m_{11}+\Pr(y^C{}_2|s)\cdot m_{21}+\Pr(y^C{}_3|s)\cdot m_{31}+..+\Pr(y^C{}_{|Y_C|}|s)\cdot m_{|Y_C|,1}\}$

But, based on the definition of ***M****, the expression above is actually equal to $\Pr(y^A{}_1|s)\cdot \Pr(y^B{}_1|s)$, which, due to the orthogonality of ***A*** and ***B***, is equal to the value in the first row, first column of $\boldsymbol{G}^{AB}$.

◙

Theorem 4.5.2.1 (Monotonicity). ***A***, ***B***, ***C***, ***D***, are ISs defined on $S\text{x}Y_A$, $S\text{x}Y_B$, $S\text{x}Y_C$, and $S\text{x}Y_D$, respectively. ***A*** and ***B*** are orthogonal, and ***C*** and ***D*** are orthogonal. Let $\boldsymbol{G}^{AB}$ denote the integration of ***A*** and ***B*** ($\boldsymbol{G}^{AB}$=***A***@***B***). Let $\boldsymbol{G}^{CD}$ denote the integration of ***C*** and ***D*** ($\boldsymbol{G}^{CD}$=***C***@***D***). Then the relationship between the informativeness of ***A***, ***B***, ***C***, ***D***, and the informativeness of $\boldsymbol{G}^{CD}$ and $\boldsymbol{G}^{AB}$ is monotonic.

Proof:

Let ***A'*** denote an IS defined on $S\text{x}Y_{A'}$ as follows: ***A'*** has the same matrix as ***A***, and ***A'*** and ***D*** are orthogonal. Consider $\boldsymbol{G}^{AB}$=***A***@***B***, and $\boldsymbol{G}^{A'D}$=***A'***@***D***. The conditions of Lemma 4.5.2.1 are met, so ***D*** is GMI than ***B*** implies $\boldsymbol{G}^{A'D}$ is GMI than $\boldsymbol{G}^{AB}$.

Consider $\boldsymbol{G}^{CD}$=***C***@***D***, and $\boldsymbol{G}^{A'D}$=***A'***@***D***. The conditions of Lemma 4.5.2.1 apply, so ***C*** is GMI than ***A'*** implies $\boldsymbol{G}^{CD}$ is GMI than $\boldsymbol{G}^{A'D}$.

Together, $\boldsymbol{D}$ is GMI than $\boldsymbol{B}$ and $\boldsymbol{C}$ is GMI than $\boldsymbol{A'}$ imply $\boldsymbol{G}^{CD}$ is GMI than $\boldsymbol{G}^{A'D}$ and $\boldsymbol{G}^{A'D}$ is GMI than $\boldsymbol{G}^{AB}$. Due to the transitivity of the ranking by informativeness, $\boldsymbol{D}$ is GMI than $\boldsymbol{B}$ and $\boldsymbol{C}$ is GMI than $\boldsymbol{A'}$ imply $\boldsymbol{G}^{CD}$ is GMI than $\boldsymbol{G}^{AB}$.

It is easy to see that $\boldsymbol{A'}$ is informatively equivalent to $\boldsymbol{A}$ (i.e, $\boldsymbol{A}$ is GMI than $\boldsymbol{A'}$ and vice versa). Therefore, the above can be re-stated: $\boldsymbol{D}$ is GMI than $\boldsymbol{B}$ and $\boldsymbol{C}$ is GMI than $\boldsymbol{A}$ imply $\boldsymbol{G}^{CD}$ is GMI than $\boldsymbol{G}^{AB}$.

◙

<u>Example 4.5.2.1:</u> Suppose that an IS $\boldsymbol{A}$ is the same as IS 4.8, $\boldsymbol{B}$ is the same as IS 4.9, and $\boldsymbol{A}$ and $\boldsymbol{B}$ are orthogonal. Suppose also that $\boldsymbol{C}$ has the same matrix as IS 4.8, $\boldsymbol{D}$ has the same matrix as IS 4.9, and $\boldsymbol{C}$ and $\boldsymbol{D}$ are orthogonal. It was shown (see matrix 4.10) that $\boldsymbol{D}$ is GMI than $\boldsymbol{B}$. $\boldsymbol{C}$ is GMI than $\boldsymbol{A}$ by the reflexivity of the ranking by informativeness.

Now, compare $\boldsymbol{G}^{AB}=\boldsymbol{A}@\boldsymbol{B}$ to $\boldsymbol{G}^{CD}=\boldsymbol{C}@\boldsymbol{D}$. The conditions of Theorem 4.5.2.1 are met: $\boldsymbol{C}$ and $\boldsymbol{D}$ are orthogonal, $\boldsymbol{A}$ and $\boldsymbol{B}$ are orthogonal. Since $\boldsymbol{C}$ is GMI than $\boldsymbol{A}$, and $\boldsymbol{D}$ is GMI than $\boldsymbol{B}$, Theorem 4.5.2.1 predicts that $\boldsymbol{G}^{CD}$ is GMI than $\boldsymbol{G}^{AB}$. This prediction will be illustrated next.

$\boldsymbol{G}^{AB}$ is given by IS 4.11.

$\boldsymbol{G}^{CD}$ is the following:

(4.19)

Signal /State	y^{CD}_1	y^{CD}_2	y^{CD}_3	y^{CD}_4
s_1	0.81	0.09	0.09	0.01
s_2	0.36	0.24	0.24	0.16

$\boldsymbol{G}^{CD}$ is GMI than $\boldsymbol{G}^{AB}$. Especially, a Markov matrix $\boldsymbol{M}$ such that $\boldsymbol{G}^{CD}\boldsymbol{M}$=$\boldsymbol{G}^{AB}$ is derived from matrix 4.10 as explained in the proof of Lemma 4.5.2.1, by duplicating it along the diagonal; the remaining values are set to zero:

(4.20)

Signal /Signal	y^{AB}_1	y^{AB}_2	y^{AB}_3	y^{AB}_4
y^{CD}_1	.6	.4	0	0
y^{CD}_2	.3	.7	0	0
y^{CD}_3	0	0	.6	.4
y^{CD}_4	0	0	.3	.7

◘

4.5.3 Perfect information is not achievable

Monotonicity was proved based on the assumption of orthogonality, regardless of whether the ISs are GMI-related or not. However, if, in addition to orthogonality, the ISs are GMI-related, then their integration has an upper-boundary on its informativeness, which is strictly inferior to perfect IS. Theorem 4.5.3.1 and Theorem 4.5.3.2 prove this. A potentially useful property of the proposed boundary is that it is calculable from the IS that has the highest ranking among the input ISs.

Theorem 4.5.3.1 (upper boundary). Let $\boldsymbol{A}$, $\boldsymbol{A'}$, $\boldsymbol{B}$, be ISs defined on SxY_{A}, $SxY_{\mathrm{A'}}$, and SxY_{B}, respectively, such that:

1. $\boldsymbol{A}$ and $\boldsymbol{A'}$ have equal matrices
2. $\boldsymbol{A}$ is GMI than $\boldsymbol{B}$
3. $\boldsymbol{A}$ and $\boldsymbol{B}$ are orthogonal
4. $\boldsymbol{A}$ and $\boldsymbol{A'}$ are orthogonal

Let $\boldsymbol{G}^{AB}$ denote the integration of $\boldsymbol{A}$ and $\boldsymbol{B}$ ($\boldsymbol{G}^{AB}$=$\boldsymbol{A}$@$\boldsymbol{B}$). Let $\boldsymbol{G}^{A'A}$ denote the integration of $\boldsymbol{A}$ and $\boldsymbol{A'}$ ($\boldsymbol{G}^{A'A}$=$\boldsymbol{A'}$@$\boldsymbol{A}$). Then $\boldsymbol{G}^{A'A}$ is GMI than $\boldsymbol{G}^{AB}$.

Proof:

By assumption, $\boldsymbol{A}$ and $\boldsymbol{B}$ are orthogonal, and $\boldsymbol{A}$ and $\boldsymbol{A'}$ are orthogonal. $\boldsymbol{A}$ is GMI than $\boldsymbol{B}$ by assumption, and $\boldsymbol{A'}$ is GMI than $\boldsymbol{A}$ since $\boldsymbol{A}$ and $\boldsymbol{A'}$ since are informatively equivalent. Therefore the conditions of the Monotonicity Theorem (4.5.2.1) are satisfied, such that $\boldsymbol{G}^{A'A}$ is GMI than $\boldsymbol{G}^{AB}$.

◙

Theorem 4.5.3.2 (perfect information is not achievable). Let $\boldsymbol{A}$, $\boldsymbol{A'}$, be ISs defined on SxY_{A}, $SxY_{\mathrm{A'}}$, respectively, such that $\boldsymbol{A}$ is not a perfect IS, $\boldsymbol{A}$ and $\boldsymbol{A'}$ have equal matrices, and $\boldsymbol{A}$ and $\boldsymbol{A'}$ are orthogonal. Then, perfect information is not achievable by $\boldsymbol{A}$ and $\boldsymbol{A'}$.

Proof:

Based on Lemma 4.3.1.1, a perfect IS is such if and only if it has a special structure, i.e., at most one non-zero value in each column of the matrix. By assumption, $\boldsymbol{A}$ is not a perfect IS. It follows from the lemma that there is at least one column of $\boldsymbol{A}$ with two or more non-zero values. It can be assumed—without loss of generality—that the lth column of $\boldsymbol{A}$ has two positive values, and these positive values appear in rows k and j.

Since $\boldsymbol{A}$ and $\boldsymbol{A'}$, have the same matrix, that the lth column of $\boldsymbol{A'}$ has two positive values, and these positive values appear in rows k and j.

Based on the definition of integration, and the orthogonality of $\boldsymbol{A}$ and $\boldsymbol{A'}$, the $l \cdot |Y_{\mathrm{A}}|+l$ column of $\boldsymbol{G}^{A'A}=\boldsymbol{A'}@\boldsymbol{A}$ must have two non-zero values in rows k and j. In this case, as proved by Lemma 4.3.1.1, $\boldsymbol{G}^{A'A}$ is not a perfect IS.

◙

The findings of Theorem 4.5.3.1 and Theorem 4.5.3.2 are demonstrated through Example 4.5.2.1 in the previous section. In that example, $\boldsymbol{A}$ is GMI than $\boldsymbol{B}$, and $\boldsymbol{C}$ is GMI than $\boldsymbol{D}$. Theorem 4.5.3.1 and Theorem 4.5.3.2 predict that the informativeness of $\boldsymbol{G}^{AB}$ is bounded by the informativeness of $\boldsymbol{G}^{CD}$, which, in turn, is strictly inferior to that of a perfect IS. The calculations of $\boldsymbol{G}^{AB}$ and $\boldsymbol{G}^{CD}$ agree with this prediction. $\boldsymbol{G}^{CD}$ is strictly inferior to a perfect IS (Lemma 4.3.1.1), and $\boldsymbol{G}^{AB}$ is inferior to $\boldsymbol{G}^{CD}$.

4.5.4 Summary of results: Sources are state-conditionally independent and can be ordered by their quality

Section 4.5 analyzed the informativeness and value of integration when the ISs are orthogonal, and one has higher informativeness than the other.

The analysis found that the informativeness of an integration is strictly higher than the informativeness of any of the input ISs, except for integrations in which for every pair of states, the matching sub-matrix of the more informative IS describes a perfect IS, and/or, the sub-matrix of the less informative IS describes a null IS. Intuitively, these integrations reflect "bad" complementarities among information sources, where one information source may have perfect information within some value ranges (an expert), and in value ranges on which that source does not have perfect information, the other source has no information at all.

A second finding is a monotonic relationship between the informativeness of the ISs and the informativeness of their integration. When the informativeness of the ISs grows, so does the informativeness of their integration. This outcome points to a source selection strategy in integrations under the assumed conditions: select the sources with the highest individual informativeness.

A third finding is that the informativeness of integration never reaches perfect information. Furthermore, when none of the ISs is a perfect IS, their integration has an upper-boundary on its informativeness that is strictly inferior to a perfect IS.

The table below summarizes the theoretical findings in Section 4.5.

	ISs are orthogonal	ISs are *not* orthogonal
One IS is more informative (or ISs equivalent)	-INCREASE IS NOT GUARANTEED - MONOTONICITY - PERFECT INFORMATION IS NOT ACHIEVABLE	
None is more informative		

4.6 INTEGRATION WHEN SOURCES ARE STATE-CONDITIONALLY INDEPENDENT, AND CANNOT BE ORDERED BY THEIR QUALITY

The importance of diversity in the information sources to be integrated is often justified based on analyses of the dependence relationship between information sources. For example, research on Condorcet's Jury Theorem in political science points to the value of “opposing schools of thought” in democratic societies, based on findings that negative correlations between information sources improve the accuracy of a majority vote (Ladha, 1992; Ladha 1995). Another example, research on ensemble learning algorithms in computer science applies various measures of dependence in a debate about the usefulness of dependence / independence for enhancing the classification accuracy of the algorithms (e.g., Ali and Pazzani, 1996; Kuncheva and Whitaker, 2003).

Another type of diversity emerges in this analysis. Ranking based on Blackwell sufficiency is, in general, not total. ISs that are not ranked may each be more informative than the other within some subset of the states, i.e., have complementary informative strenghts. Theorem 4.6.1.1 and the examples that follow it suggest, in agreement with common intuition, that such diversity can be very useful to information integration.

4.6.1 Increase in quality is guaranteed

Theorem 4.6.1.1 asserts the strictly higher informativeness of integrations of ISs that are not GMI-related.

Theorem 4.6.1.1. ***A*** and ***B*** are two ISs defined on SxY_A and SxY_B, respectively. If ***A*** is not GMI than ***B*** and ***B*** is not GMI than ***A***, then an increase in informativeness is guaranteed for ***A*** and ***B***.

Proof:

Theorem 4.4.1 ensures that ***G***=***A***@***B*** is GMI than both ***A*** and ***B***. Suppose that, converse to this theorem's conclusion, ***G*** is not GSMI than ***A***, that is, ***A*** is GMI than ***G***. But now the contradiction follows immediately. If ***A*** is GMI than ***G*** and ***G*** is GMI than ***B***, then, since GMI is a transitive relation, ***A*** must be GMI than ***B***, in contradiction to the assumption of the theorem.

A symmetric logic applies for the ***B***. Therefore ***G*** is GSMI than ***A*** and ***B***, such that, by definition, an increase in informativeness is guaranteed for ***A*** and ***B***.

◙

4.6.2 Monotonicity

Theorem 4.5.2.1 holds whenever the ISs taking part in the integration are orthogonal, such that it is still valid when these ISs are not GMI-related.

Theorem. $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$, $\boldsymbol{D}$, are ISs defined on SxY_{A}, SxY_{B}, SxY_{C}, and SxY_{D}, respectively. $\boldsymbol{A}$ and $\boldsymbol{B}$ are orthogonal, and $\boldsymbol{C}$ and $\boldsymbol{D}$ are orthogonal. Let $\boldsymbol{G}^{AB}$ denote the integration of $\boldsymbol{A}$ and $\boldsymbol{B}$ ($\boldsymbol{G}^{AB}$=$\boldsymbol{A}$@$\boldsymbol{B}$). Let $\boldsymbol{G}^{CD}$ denote the integration of $\boldsymbol{C}$ and $\boldsymbol{D}$ ($\boldsymbol{G}^{CD}$=$\boldsymbol{C}$@$\boldsymbol{D}$). Then the relationship between the informativeness of $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$, $\boldsymbol{D}$, and the informativeness of $\boldsymbol{G}^{CD}$ and $\boldsymbol{G}^{AB}$ is monotonic.

Proof:

See Theorem 4.5.2.1.

◙

4.6.3 Perfect information is achievable

Unlike the integration of ISs that are ranked by their informativeness, integration of ISs that are not ranked can result in a perfect IS. This will be proved by an example—an adaptation of an example from the domain of ensemble learning with neural networks (Hansen and Salamon, 1990). The example also illustrates, informally though, the specific nature of the ISs and the relationship between them that are needed for generating a perfect IS. Notably, it applies ISs that were already recognized (Theorem 4.5.1.1). An IS for which there exists a proper subset of the states such that the sub-matrix is a perfect IS, can actually produce perfect information if integrated with the "right" IS. Such IS is of the same type, which, nonetheless, has a complementary structure, i.e., for any subset of the states such that the matching sub-matrix of one of them is not a perfect IS, the sub-matrix of the other is a perfect IS. Integrations of this kind may be described intuitively as integrations of information sources that have expertise in different, complementary areas.

A closely related example is presented in Section 4.7.3. Together, the two examples may shed some light on the conditions in which perfect information is achieved.

Example 4.6.3.1: This example portrays the integration of three ISs, corresponding to three neural networks. The number of ISs (three) was left as specified in the original example, partly in order to hint to the way in which the theory in this study—which assumes two ISs—can be generalized to $n \geq 2$ ISs.

The neural networks classify the location of an object in terms of four possible location categories, I, II, III, and IV. In agreement, the state set is defined as S={I, II, III, IV}. The signal set, common to all three ISs, includes two signals, denoted **W**, and **B**. The ISs are given under (4.21)-(4.23):

(4.21)

Signal /State	W	B
I	0	1
II	1	0
III	1	0
IV	1	0

(4.22)

Signal /State	W	B
I	1	0
II	1	0
III	1	0
IV	0	1

(4.23)

Signal /State	W	B
I	0	1
II	1	0
III	0	1
IV	0	1

ISs (4.21)-(4.23) are orthogonal—the signals clearly depend only on the given state.

In addition, it can be shown that none of the ISs is GMI than the other. For example, let IS (4.21) be denoted ***A***; IS (4.22) - ***B***. Here is a proof, by contradiction, that ***A*** is not GMI than ***B***:

Suppose that ***A*** is GMI than ***B***. Therefore, there is a Markov matrix ***M*** such that ***AM***=***B***. The dimensions of ***A*** and ***B*** dictate that ***M*** has two rows and two columns. Based on the values in the first row of ***A*** and the first row of ***B***, m_{21}, the value in the second row, first column, of ***M***, must equal one. Moreover, based on the values in the second rows of ***A*** and ***B***, m_{11} must equal one. But then the other values ***M*** should be zero since ***M*** is a Markov matrix. It follows from the calculated values of ***M*** that b_{42}, the value in the fourth row, second column, of ***B***, which is derived from the fourth row of ***A*** and the second column of ***M***, is equal zero. But this contradicts the actual value in the fourth row, second column, of ***B***, which is one. Therefore ***M*** as required does not exist.

In a similar way it can be proved that ISs (4.21)-(4.23) fit the conditions discussed in this section. The integration of ***A*** and ***B*** is:

(4.24)

Signal /State	"II/III"	"IV"	"I"	-
I	0	0	1	0
II	1	0	0	0
III	1	0	0	0
IV	0	1	0	0

A second integration, this time of (4.24) and (4.23), yields:

(4.25)

Signal /State	"II"	"III"	-	"IV"	-	"I"	-	-
I	0	0	0	0	0	1	0	0
II	1	0	0	0	0	0	0	0
III	0	1	0	0	0	0	0	0
IV	0	0	0	1	0	0	0	0

None of the ISs (4.21)-(4.23) suits the conditions of Theorem 4.5.1.1. However, the ISs are complementary in the "good" sense described above. IS 4.2.1 and IS 4.2.2, are complementary over {I, II}, {I, III}, and {III, IV}; IS 4.2.2 and IS 4.2.3 are complementary over {I, II}, {II, III}, and {III, IV}; IS 4.2.1 and IS 4.2.3 are complementary over {I, III}, {I, IV}, and {II, IV}. IS (4.25)—the integration of all three ISs—is a perfect IS (Lemma 4.3.1.1).◘

4.6.4 Summary of results: Sources are state-conditionally independent and cannot be ordered by their quality

Section 4.6 analyzed the informativeness and value of integration when the ISs are orthogonal and none has higher informativeness than the other.

The analysis found that the informativeness of the integration is strictly higher than the informativeness of any of the input ISs.

A second finding is a monotonic relationship between the informativeness of the ISs and the informativeness of their integration.

A third finding is that some integrations reach perfect information quality. An example identifies these integrations. An IS for which there exists a proper subset of the states such that the sub-matrix is a perfect IS, can create perfect information if integrated with an IS that is of the same type, and has a complementary structure. For any subset of the states such that the sub-matrix of one of them is not a perfect IS, the matching sub-matrix of the other is a perfect IS. Integrations of this kind may be described as information sources that have expertise in different, complementary areas. The relevant ISs were noted by earlier analysis (Theorem 4.5.1.1). Evidently, ISs in this category are particularly sensitive to the choice of ISs with which they are integrated. A worst case scenario is zero increase in informativeness. A best case scenario is perfect information.

The results in this section, especially the findings in regard to perfect information, indicate the potential value of a strategy that seeks complementary informative strengths among information sources. The theory in this section implies that even when an information source is not highly informative as a whole, “slices” (subsets) of its overall value range may still be highly informative. When different information sources have complementary "slices" of this kind, their integration can, under the best scenario, reach perfect information quality and value.

The diagram below summarizes the theoretical findings in this section.

	ISs are orthogonal	ISs are *not* orthogonal
One IS is more informative (or ISs equivalent)		
None is more informative	-Increase is guaranteed -Monotonicity - Perfect information is achievable	

4.7 Integration when sources are not state-conditionally independent, and can be ordered by their quality

4.7.1 Increase in quality is not guaranteed

The integrations of ISs that are not orthogonal, and one has higher informativeness than the other, are not always strictly more informative than the input ISs.

When one of the ISs ***A***, ***B***, is GMI than the other—say, ***A*** is GMI than ***B***—then, by definition, there is a Markov matrix ***M*** such that ***AM***=***B***. ***M*** determines the values of another Markov matrix ***M****, such that ***AM**** can happen to be equal to the integration of ***A*** and ***B***. The actual outcome depends on the specifics of the state-conditional dependence between ***A*** and ***B***. Yet, if ***G***, the integration of ***A*** and ***B***, is indeed equal to ***AM****, then ***A*** is GMI than ***G*** by definition.

Fortunately, when ISs ***A***, ***B***, are not orthogonal and ***A*** is GMI than ***B***, ***A*** is GMI than ***G*** only if ***G*** is equal to ***AM****.

Figures A.1 and A.2 in the Appendix portray ***M****. Figure 4.1 and depict the "problematic" integration. The lowercase terms in Figure 4.2 point to elements of ***A*** and ***M***. For example, $a_{11}m_{11}$ is a multiple of the value in the first row, first column of ***A***, and the value in the first row, first column of ***M***.

Theorem 4.7.1.1 formalizes the described understanding.

FIGURE 4.1: Structure of integration when there is no increase in informativeness

COLUMNS ARE LINEARLY DEPENDENT ON THE 1ST COLUMN OF ***A***	COLUMNS ARE LINEARLY DEPENDENT ON THE 2ND COLUMN OF ***A***	……. …….	……. …….	COLUMNS ARE LINEARLY DEPENDENT ON THE $\lvert Y_A \rvert$th COLUMN OF ***A***
$\lvert Y_B \rvert$ COLUMNS	$\lvert Y_B \rvert$ COLUMNS	…….	…….	$\lvert Y_B \rvert$ COLUMNS

FIGURE 4.2: Structure of integration when there is no increase (refined view)

$$
\begin{array}{ccccccccc}
a_{1,1}\boldsymbol{m}_{1,1} & a_{1,1}\boldsymbol{m}_{1,2} & .. & a_{1,1}\boldsymbol{m}_{1,c} & a_{1,2}\boldsymbol{m}_{2,1} & a_{1,2}\boldsymbol{m}_{2,2} & .. & a_{1,2}\boldsymbol{m}_{2,c} & \\
a_{2,1}\boldsymbol{m}_{1,1} & a_{2,1}\boldsymbol{m}_{1,2} & .. & a_{2,1}\boldsymbol{m}_{1,c} & a_{2,2}\boldsymbol{m}_{2,1} & a_{2,2}\boldsymbol{m}_{2,2} & .. & a_{2,2}\boldsymbol{m}_{2,c} & \\
a_{3,1}\boldsymbol{m}_{1,1} & a_{3,1}\boldsymbol{m}_{1,2} & .. & a_{3,1}\boldsymbol{m}_{1,c} & a_{3,2}\boldsymbol{m}_{2,1} & a_{3,2}\boldsymbol{m}_{2,2} & .. & a_{3,2}\boldsymbol{m}_{2,c} & \\
.. & & & & .. & & & & \\
.. & & & & .. & & & & \bullet\ \bullet\ \bullet \\
.. & & & & .. & & & & \\
.. & & & & .. & & & & \\
a_{r,1}\boldsymbol{m}_{1,1} & a_{r,1}\boldsymbol{m}_{1,2} & \bullet & a_{r,1}\boldsymbol{m}_{1,c} & a_{r,2}\boldsymbol{m}_{2,1} & a_{r,2}\boldsymbol{m}_{2,2} & .. & a_{r,2}\boldsymbol{m}_{2,c} &
\end{array}
$$

Theorem 4.7.1.1. ***A***, ***B***, are ISs defined on SxY_{A}, SxY_{B}, respectively, such that ***A*** is GMI than ***B***. Let ***M*** denote a Markov matrix such that ***AM***=***B***. Let ***G*** denote the integration of ***A*** and ***B*** (***G***=***A***@***B***). Then, ***A*** is GMI than ***G*** if and only if for i=1,..,$|Y_{\mathrm{A}}|$, each of the columns in the ith sequence of $|Y_{\mathrm{B}}|$ columns of ***G*** is equal to a scalar multiplication of the ith column of ***A*** by the values, each in turn, of the ith row of ***M***.

Proof:

a_{ij} will denote the value in row i column j of ***A***; b_{ij} will denote the value in row i column j of ***A***. g_{ij} will denote the value in row i column j of ***G***. m_{ij} will denote the value in row i column j of ***M***.

(⇐)

Suppose that ***G*** can be described by the values of ***A*** and ***M*** according to the pattern assumed by this theorem. Therefore, the values of a matrix ***M**** can be derived from the values of ***G*** as follows.

M* has $|Y_A|$ rows, and $|Y_A \times Y_B|$ columns. The values of ***M**** are zero, except for one-row blocks along its diagonal, consisting of $|Y_B|$ columns each. A block is derived from the respective columns of ***G***. The first block in ***M****, appearing in its first row, is derived from the first $|Y_B|$ columns of ***G***—the values are $m_{1,1}, m_{1,2}, \ldots, \ldots m_{1,|Y_B|}$, i.e., the scalars that multiply the first column of ***A***. The second block, which appears in the second row of ***M**** beginning in column $|Y_B|+1$ in that row, shows $m_{2,1}, m_{2,2}, \ldots\ m_{2,|Y_B|}$, i.e., the scalars that multiply the second column of ***A***. etc. Altogether there are $|Y_A|$ blocks of this kind, one in each row of ***M****, and consistent with the number of columns of ***G***, which is, by definition, $|Y_A \times Y_B|$.

M* is a Markov matrix: The values in each row of ***M**** are all zero except for $|Y_B|$ values along the diagonal, which are the same as the values in the respective row in a Markov matrix, ***M***.

AM*=***G***. ***M**** has the required dimensions for such product: The number of rows of ***M****, $|Y_A|$, is the same as the number of columns of ***A***; the number of columns is $|Y_A \times Y_B|$, the same as the number of columns of ***G***.

Consider the product of the first row of ***A*** and first column of ***M****. By the definition of ***M**** all the values in its first column are zero, except for the value in the first row which is equal $m_{1,1}$. Therefore, such product has the form:

$a_{1,1}\, m_{1,1} + a_{1,2}\cdot 0 + \ldots + a_{1,|Y_A|}\cdot 0 = a_{1,1}\, m_{1,1}$

But $a_{1,1}\, m_{1,1}$ is, by assumption, the value of $g_{1,1}$.

A similar calculation for any row of ***A*** and column of ***M**** will show that ***AM****=***G***.

Therefore, ***A*** is GMI than ***G***.

(⇒)

Suppose that ***A*** is GMI than ***G***.

If ***A*** is GMI than ***G***, there is a Markov matrix ***M**** such that ***AM*=G***. According to the proof of Theorem 4.5.1.1, there are, consequently, linear dependence constraints on ***G***. Specifically, there exist values $k_{i,j}$ i=1..$|Y_A|$, j=1.. $|Y_B|$ such that[3]:

$g_{1,1}=a_{1,1}\cap b_{1,1}=k_{1,1}a_{1,1}$; $g_{2,1}=a_{2,1}\cap b_{2,1}=k_{1,1}a_{2,1}$; … $g_{|S|,1}=a_{|S|,1}\cap b_{|S|,1}=k_{1,1}a_{|S|,1}$

$g_{1,2}=a_{1,1}\cap b_{1,2}=k_{1,2}a_{1,1}$; $g_{2,2}=a_{2,1}\cap b_{2,2}=k_{1,2}a_{2,1}$; … $g_{|S|,2}=a_{|S|,1}\cap b_{|S|,2}=k_{1,2}a_{|S|,1}$

…

$g_{1,|Y_B|}=a_{1,1}\cap b_{1,|Y_B|}=k_{1,|Y_B|}a_{1,1}$; $g_{2,|Y_B|}=a_{2,1}\cap b_{2,|Y_B|}=k_{1,|Y_B|}a_{2,1}$; … $g_{|S|,|Y_B|}=a_{|S|,1}\cap b_{|S|,|Y_B|}=$

$k_{1,|Y_B|}a_{|S|,1}$

[3] To keep notation short I compromise on consistency. The notation $a_{1,1}$, $b_{1,1}$,.. for values of ***A*** and ***B*** designates conditional probabilities. The notation of the values of ***G***, e.g., $a_{1,1}\cap b_{1,1}$ designate the respective joint conditional probabilities.

$g_{1,|Y_B|+1}=a_{1,2}\cap b_{1,1}=k_{2,1}a_{1,2}$; $g_{2,|Y_B|+1}=a_{2,2}\cap b_{2,1}=k_{2,1}a_{2,2}$; … $g_{|S|,|Y_B|+1}=a_{|S|,2}\cap b_{|S|,1}=k_{2,1}a_{|S|,2}$

…

Therefore, let the values of ***M*** be defined in the following manner:

$m_{i,j,}\equiv k_{i,j} \qquad i=1..|Y_A|, j=1..|Y_B|$

$k_{i,j}$ are non-negative since ***G*** and ***A*** are Markov Matrices. Moreover, due to the linear constraints as above, $a_{l,i}\cdot\sum_j k_{i,j}=a_{l,i}$ for any valid l,i. Therefore, $\sum_j k_{i,j}=1$. Therefore, ***M*** is a Markov matrix.

AM=***B***, so ***A*** is GMI than ***B***. First, ***M*** has the required dimensions for such product, i.e., $|Y_A|$ rows and $|Y_B|$ columns. Furthermore, the product of the first row of ***A*** and the first column of ***M***, for example, is:

$a_{1,1}\, k_{1,1} + a_{1,2}\cdot k_{2,1} + \ldots + a_{1,|Y_A|}\cdot k_{|Y_A|,1} = a_{1,1}\cap b_{1,1} + a_{1,2}\cap b_{1,1} + .. \, a_{1,|Y_A|}\cap b_{1,1} = b_{1,1}$

A similar calculation for any row of ***A*** and column of ***M*** will show that ***AM***=***B***.

The constraints also show that ***G*** is derived from ***A*** and ***M*** as specified by this theorem.

◙

Example 4.7.1.1: To illustrate Theorem 4.7.1.1, consider IS ***A*** (4.8) and ***B*** (4.9). It was demonstrated (4.10) that ***AM***=***B***, i.e., ***A*** is GMI than ***B***. Suppose, however, that, unlike earlier assumption, ***A*** and ***B*** are not orthogonal. More accurately, suppose that their integration is derived from ***M*** (4.10) and ***A*** in the manner described by Theorem 4.7.1.1, i.e. , $\boldsymbol{G}^{AB}=\boldsymbol{A}@\boldsymbol{B}$ is:

(4.26)

Signal /State	y^{AB}_1	y^{AB}_2	y^{AB}_3	y^{AB}_4
s_1	0.54	0.36	0.03	0.07
s_2	0.36	0.24	0.12	0.28

A is GMI than $\boldsymbol{G}^{AB}$. To show it, a Markov matrix ***M**** such that $\boldsymbol{AM^*}=\boldsymbol{G}^{AB}$ is derived from (4.10) in the manner described by the proof of Theorem 4.7.1.1:

(4.27)

Signal /Signal	y^{AB}_1	y^{AB}_2	y^{AB}_3	y^{AB}_4
y^{A}_1	0.6	0.4	0	0
y^{A}_2	0	0	0.3	0.7

Therefore, if the conditional joint distributions of ***A*** (4.8) and ***B*** (4.9) are the same as specified under (4.26), there is no increase in informativeness for ***A*** and ***B***.

◘

4.7.2 No monotonicity; perfect information is achievable

The following example proves both that perfect information is achievable in the set of integrations that satisfy the assumed conditions in this section, and that the relationship between the informativeness of the ISs and the informativeness of their integration is not monotonic there.

Example 4.7.2.1: Consider two ISs, ***A*** (4.28, left) and ***B*** (4.28, right). It can be shown that ***A*** is GMI than ***B***, in agreement with the assumptions in this section. It was already proved

(Theorem 4.5.1.1) that if such ISs are orthogonal, their integration is not a perfect IS (this can be immediately verified). However, can any possible integration of ***A*** and ***B*** be a perfect IS (i.e., when ***A*** and ***B*** are not orthogonal)?

(4.28)

Signal /State	$y^A{}_1$	$y^A{}_2$
s_1	.9	.1
s_2	.2	.8

Signal /State	$y^B{}_1$	$y^B{}_2$
s_1	.64	.36
s_2	.92	.08

It will be proved that no-matter their specific dependence relationship, the integration of ***A*** and ***B*** will not be a perfect IS. Subsequently, ***A*** and ***B*** will be compared with other ISs in this quadrant whose integration may reach perfect information quality, depending on their specific dependence relationship.

Let $\boldsymbol{G}^{AB}$ denote an integration of ***A*** and ***B***. Consider $g_{11}=\Pr(y^A{}_1 \cap y^B{}_1 | s_1)$, the value in the first row, first column of $\boldsymbol{G}^{AB}$. g_{11} can only get values in the range 0.54-0.64. If g_{11} is higher than 0.64, then $g_{11}=\Pr(y^A{}_1 \cap y^B{}_1 | s_1) > \Pr(y^B{}_1 | s_1)$, contrary to a definition of probability. If g_{11} is lower than 0.54, then, since $g_{11}+g_{12}$ must be equal to $\Pr(y^A{}_1 | s_1)=0.9$, $g_{12}=\Pr(y^A{}_1 \cap y^B{}_2 | s_1)>0.36$, i.e., $g_{12}=\Pr(y^A{}_1 \cap y^B{}_2 | s_1) > \Pr(y^B{}_2 | s_1)$, which, again, contradict the definition of probability.

Next, consider $g_{21}=\Pr(y^A{}_1 \cap y^B{}_1 \ | s_2)$, the value in the second row, first column of of $\boldsymbol{G}^{AB}$. Similar to g_{11}, g_{21} can only have values in the range 0.12-0.2. Therefore, both g_{11} and g_{21} are positive. Therefore $\boldsymbol{G}^{AB}$ is not a perfect IS (Lemma 4.3.1.1), no matter the specific dependence relationship between ***A*** and ***B***.

Consider now another IS, ***C***, which is given by:

(4.29)

Signal /State	y^{C}_{1}	y^{C}_{1}
s_1	0.64	0.36
s_2	0.5	0.5

It can be shown that both ***A*** and ***B*** are GMI than ***C***.

Suppose that ***C*** is integrated with an IS ***C′*** that has the same matrix as ***C***, and ***C***, ***C′***, are strongly conditionally dependent, i.e., ***C***, ***C′***, are GMI-related (equivalent) and are not orthogonal, in agreement with the assumptions in this section. In particular, suppose that their integration $\boldsymbol{G}^{CC}$=***C***@***C′*** is:

(4.30)

Signal /State	$y^{CC'}_{1}$	$y^{CC'}_{2}$	$y^{CC'}_{3}$	$y^{CC'}_{4}$
s_1	0.64	0	0	0.36
s_2	0	0.5	0.5	0

$\boldsymbol{G}^{CC}$ is a perfect IS (Lemma 4.3.1.1). Therefore, despite the fact that both ***A*** and ***B*** are GMI than ***C*** and ***C′***, the integration of ***C*** and ***C′*** is a perfect IS, while no integration of ***A*** and ***B*** is a perfect IS. Hence no monotonicity, and perfect information is achievable.

◘

4.7.3 Perfect information is achievable—a second look

The example in Section 4.7.2 proved that perfect information is achievable in the set of integrations defined on *S* where the input ISs are not orthogonal, and one is GMI than the other. This section investigates the situation where perfect information is achievable by ***A*** and ***B*** in more detail, first through a variation on Example 4.6.3.1.

Example 4.7.3.1: The ISs model the same neural networks that were represented by (4.6.1) and (4.6.2)[4] as orthogonal, non-GMI-related ISs.

The neural networks' task is assumed to be more simple than in Example 4.6.3.1. In particular, category I and category IV are taken as one state, and so are categories II and III, such that the state set consists of I/IV, and II/III. The signal set includes, as before, **W**, and **B**.

IS (4.21) translates to IS (4.31):[5]

(4.31)

Signal /State	W	B
I/IV	.5	.5
II/III	1	0

Similarly, IS (4.22) translates to IS (4.32):

[4] The third neural network was omitted from the example partly in order to keep the analysis consistent with the assumptions of this section. This choice does not affect the conclusions.
[5] The prior probabilities of the quadrants are taken to be (.25, .25, .25, .25).

(4.32)

Signal /State	W	B
I/IV	.5	.5
II/III	1	0

Note that (4.32) is identical to (4.31), so that the ISs are informatively equivalent, i.e., GMI-related. Their integration, consistent with (4.24), is the following:

(4.33)

Signal /State	"II/III"	"I/IV"	"I/IV"	-
I/IV	0	.5	.5	0
II/III	1	0	0	0

Matrix (4.33) is a perfect IS (Lemma 4.3.1.1). When matrices (4.31)-(4.33) are considered together, the values in the matrices imply that ISs (4.31) and (4.32) are not orthogonal. If they were orthogonal, the values in the first row of (4.33) were uniformly 0.25. There is a strong conditional dependence given state I/IV.

◘

Example 4.6.3.1 and Example 4.7.3.1 describe identical information sources, yet, Example 4.6.3.1 models the sources by orthogonal, non-GMI-related ISs, while the example in this section models the same sources by GMI-related, strongly conditionally dependent ISs. The contrasting representations are alternative manifestations of a situation described earlier as "information sources that have expertise in different, complementary areas". In Example 4.6.3.1 such sources are displayed with relatively high resolution, so their conditional independence shows, and their complementary nature is exposed by the lack of ranking. The example in this section shows the same sources with lower resolution, so their complementarity is revealed through the strong conditional dependence. Both forms are probably widespread in practical integration scenarios.

Theorem 4.7.3.1 generalizes Example 4.7.2.1 and Example 4.7.3.1. It investigates the general characteristics of non-orthogonal ISs whose integration is a perfect IS. The analysis focuses on a subset of the non-orthogonal ISs defined on S—it is not exhaustive. However, it may still point to some fundamental requirements for the production of perfect information. Especially, the theorem points to a strong state-conditional dependence.

Theorem 4.7.3.1 Let $\boldsymbol{A}$, $\boldsymbol{B}$, denote two ISs defined on SxY_A, SxY_B, respectively, such that $|S|=|Y_A|=|Y_B|$, and the elements of $\boldsymbol{A}$, $\boldsymbol{B}$, are all non-zero.

$\boldsymbol{G}=\boldsymbol{A}@\boldsymbol{B}$ is a perfect IS, if and only if $\boldsymbol{A}$, $\boldsymbol{B}$, have equal matrices, at least one of the state-conditional distributions is uniform, and there is a mapping f:Sx$Y_A\rightarrow Y_B$ onto Y_B, such that $\Pr(y^B|s,y^A)=1$ $\forall$ $s\in S$, $y^A\in Y_A$, $y^B=f(s,y^A)$, and $f:(s_j,y^A{}_i)\neq f:(s_k,y^A{}_i)$, $\forall$ $j\neq k$.

Proof:

The proof will be given for $\boldsymbol{A}$, $\boldsymbol{B}$, that are 2X2 ISs. A similar logic can be applied to the more general case of nXn ISs.

Let $a_{ij}=\Pr(y^A{}_j|s_i)$ denote the value in row i, column j, of $\boldsymbol{A}$; let $b_{ij}=\Pr(y^B{}_j|s_i)$ denote the value in row i, column j, of $\boldsymbol{B}$; let g_{ij} denote the value in row i, column j, of $\boldsymbol{G}$.

(⇒) The definition of probability dictates that:

1. $g_{11}+g_{12}=\Pr(y^A{}_1\cap y^B{}_1\,|s_1)+\Pr(y^A{}_1\cap y^B{}_2\,|s_1)=\Pr(y^A{}_1|s_1)=a_{11}$
2. $g_{13}+g_{14}=\Pr(y^A{}_2\cap y^B{}_1\,|s_1)+\Pr(y^A{}_2\cap y^B{}_2\,|s_1)=\Pr(y^A{}_2|s_1)=a_{12}$
3. $g_{11}+g_{13}=\Pr(y^A{}_1\cap y^B{}_1\,|s_1)+\Pr(y^A{}_2\cap y^B{}_1\,|s_1)=\Pr(y^B{}_1|s_1)=b_{11}$
4. $g_{12}+g_{14}=\Pr(y^A{}_1\cap y^B{}_2\,|s_1)+\Pr(y^A{}_2\cap y^B{}_2\,|s_1)=\Pr(y^B{}_2|s_1)=b_{12}$
5. $g_{11}+g_{12}+g_{13}+g_{14}=a_{11}+a_{12}=b_{11}+b_{12}=1$

Similarly, for the second row of ***G***:

6. $g_{21}+g_{22}=\Pr(y^A{}_1\cap y^B{}_1\,|\,s_2)+\Pr(y^A{}_1\cap y^B{}_2\,|\,s_2)=\Pr(y^A{}_1|s_2)=a_{21}$
7. $g_{23}+g_{24}=\Pr(y^A{}_2\cap y^B{}_1\,|\,s_2)+\Pr(y^A{}_2\cap y^B{}_2\,|\,s_2)=\Pr(y^A{}_2|\,s_2)=a_{22}$
8. $g_{21}+g_{23}=\Pr(y^A{}_1\cap y^B{}_1\,|\,s_2)+\Pr(y^A{}_2\cap y^B{}_1\,|\,s_2)=\Pr(y^B{}_1|\,s_2)=b_{21}$
9. $g_{22}+g_{24}=\Pr(y^A{}_1\cap y^B{}_2\,|\,s_2)+\Pr(y^A{}_2\cap y^B{}_2\,|\,s_2)=\Pr(y^B{}_2|\,s_2)=b_{22}$
10. $g_{21}+g_{22}+g_{23}+g_{24}=a_{21}+a_{22}=b_{21}+b_{22}=1$

In addition, if ***G*** is a perfect IS then each column of ***G*** can have only one positive value (Lemma 4.3.1.1). This implies that, given the positive values in, say, the first row of ***G***, the values in the second row of ***G*** in the respective columns are determined simultaneously (i.e. equal zero).

Subtraction of equation 3 from equation 1, or equation 4 from equation 2, shows that g_{12}-g_{13}=a_{11}-b_{11}. As a result, g_{12}-g_{13} must be non-zero unless a_{11}=b_{11}. Similar analysis of equations 6-10 shows that g_{22}-g_{23}=a_{21}-b_{21}. As a result, g_{22}-g_{23} must be non-zero unless a_{21}=b_{21}.

Suppose that, contrary to the conclusion of the theorem, $a_{11}\neq b_{11}$. Then, at least one of g_{12}, g_{13} , must be positive. Suppose, without loss of generality, that g_{12} is positive. Suppose also that, contrary to the conclusion of the theorem, $a_{21}\neq b_{21}$. Therefore, at least one of g_{22} , g_{23} , must be positive. However, if g_{22} is positive, then both g_{12} and g_{22} are positive, and therefore, according to Lemma 4.3.1.1, ***G*** is not a perfect IS. Suppose, therefore, that g_{22}=0, and g_{23} is positive.

If g_{13} is positive as well, then, again, ***G*** is not a perfect IS. So, g_{13}=0. Therefore, g_{12}=a_{11}-b_{11}, g_{13}=0, g_{22}=0, g_{23}=a_{21}-b_{21}. It follows from equation 1 that g_{11}=b_{11}; similarly, based on equation 6, g_{21}=a_{21}. Consequently, since both both g_{11} and g_{21} are positive, ***G*** is not a perfect IS.

Since the assumptions that $a_{11}\neq b_{11}$, $a_{21}\neq b_{21}$, did not produce the desired ***G***, consider the case that $a_{11}\neq b_{11}$, a_{21}=b_{21}. It follows, as before, that g_{12}=a_{11}-b_{11}, g_{13}=0, and, in addition, g_{22}=g_{23}. If g_{12} and g_{22} are both positive, then ***G*** is not a perfect IS. Therefore, suppose, first, that g_{22}=g_{23}=0. Then, based on equations 6 and 8, g_{21}=a_{21}=b_{21}. But, since equation 1 directs that g_{11}=b_{11}, g_{11} and g_{21} are both positive, and therefore ***G*** is not a perfect IS.

Since the assumptions that $a_{11}\neq b_{11}$, a_{21}=b_{21} led to contradiction, consider the case that a_{11}=b_{11}, a_{21}=b_{21}. (The case a_{11}=b_{11}, $a_{21}\neq b_{21}$ is symmetric to the former.) Since ***A***, ***B***, are Markov matrices, these assumptions imply that ***A***, ***B***, have equal matrices.

It was shown that $g_{12}=g_{13}$, and $g_{22}=g_{23}$. Since it cannot be that both g_{12} and g_{22} are positive, suppose, without loss of generality, that $g_{12}=g_{13}=0$. Therefore, $g_{11}=a_{11}=b_{11}$, $g_{14}=a_{12}=b_{12}$. Now, if $g_{21}\neq 0$ or $g_{24}\neq 0$, then ***G*** is not a perfect IS. Therefore, $g_{21}=g_{24}=0$. Therefore, it must be that $g_{22}=g_{23}=0.5$.

To conclude, ***G*** that is a perfect IS has the following form:

Signal /State	y^{AB}_1	y^{AB}_2	y^{AB}_3	y^{AB}_4
s_1	a_{11}	0	0	a_{12}
s_2	0	0.5	0.5	0

Which implies, based on equations 6-10, $a_{21}=b_{21}=a_{22}=b_{22}=0.5$. Therefore, ***A***, ***B***, have equal matrices, of the form:

$$\begin{vmatrix} a_{11} & a_{12} \\ 0.5 & 0.5 \end{vmatrix}$$

$f{:}S\times Y_A\rightarrow Y_B$ is defined as follows: $f{:}(s_1,\ y^A_1)=y^B_1$; $f{:}(s_1,\ y^A_2)=y^B_2$; $f{:}(s_2,\ y^A_1)=y^B_2$; $f{:}(s_2,y^A_1)=y^B_1$. It can be easily seen that $f{:}(s_j,y^A_i)\neq f{:}(s_k,y^A_i)$, $\forall\ j\neq k$, and $\Pr(y^B|s,y^A)=1\ \forall\ s\in S$, $y^A\in Y_A$, $y^B=f(s,y^A)$.

($\Leftarrow$) Suppose that ***A***, ***B***, have equal matrices, at least one of the state-conditional distributions is uniform, and there is a mapping $f{:}S\times Y_A\rightarrow Y_B$ onto Y_B, such that $\Pr(y^B|s,y^A)=1\ \forall\ s\in S$, $y^A\in Y_A$, $y^B=f(s,y^A)$, and $f{:}(s_j,y^A_i)\neq f{:}(s_k,y^A_i)$, $\forall\ j\neq k$. It can be easily shown that ***G***=***A***@***B*** has the form as illustrated in the first part of this proof, which, according to Lemma 4.3.1.1, implies a perfect IS.◘

4.7.4 Summary of results: Sources are not state-conditionally independent and can be ordered by their quality

Section 4.7 analyzed the informativeness and value of integration when the ISs are not orthogonal, and one has higher informativeness than the other.

The analysis found non-monotonic relationship between the informativeness of the ISs and the informativeness of their integration. The dependence relationship between the ISs is an additional, powerful, factor that can change the outcome up or down.

A second finding is that perfect information is achievable. An attempt to get a more detailed profile of the ISs whose integration is a perect IS shows a very different picture from that depicted in the case of orthogonal, non-GMI-related ISs. Theorem 4.7.3.1 points to the similarity of the IS matrices and a strong state-conditional dependence when their integration is a perfect IS.

An example suggests, however, that the contrasting characteristics of non-orthogonal, non-GMI-related ISs that produces perfect information versus GMI-related, orthogonal ISs that produces perfect information, may refer to the same underlying phenomenon. This phenomenon was described earlier as sources having expertise in different, complementary areas. More precisely, the contrasting characteristics make different representations of the same situation, which vary in their resolutions. The example indicates the usefulness of a tool for handling integration problems: changes in representation.

A third finding is that the informativeness of the integration is not always strictly higher than the informativeness of any of the ISs. Unlike orthogonal ISs where the ISs that may yield no increase in informativeness form a limited subset, no IS is excluded from this possibility in this quadrant. Yet, Theorem 4.7.1.1 may provide some help in prior identification of such risk. Suppose that ***M*** has been calculated, such that ***M****, as well as the "bad" integration are known. In this case the "bad" conditional dependence between ***A*** and ***B*** is known too, such that its plausibility may be assessed in advance, even based on crude understanding with respect to the state-conditional joint distributions of ***A*** and ***B***.[6]

The diagram below summarizes the theoretical findings in section 4.7.

	ISs are orthogonal	ISs are *not* orthogonal
One IS is more informative (or ISs equivalent)		- Increase not guaranteed - No monotonicity - Perfect information is achievable
None is more informative		

[6] Further study might have shown that the relationship between non-orthogonal ISs whose integration does not offer higher informativeness and orthogonal ISs whose integration does not offer higher integration is similar to the relationship uncovered in regard to perfect information. However, this possibility is not studied in this chapter.

4.8 Integration when sources are not state-conditionally independent, and cannot be ordered by their quality

The results in this section correspond to a combined effect of the conditions discussed in Section 4.6 and Section 4.7.

4.8.1 Increase in quality is guaranteed

Theorem 4.6.1.1 holds whenever the ISs are not GMI-related, no-matter the orthogonality condition. Therefore, the informativeness of the integration is strictly higher than the informativeness of the ISs under the current assumptions as well.

Theorem. ***A*** and ***B*** are two ISs defined on SxY_A and SxY_B, respectively. If ***A*** is not GMI than ***B*** and ***B*** is not GMI than ***A***, then an increase in informativeness is guaranteed for ***A*** and ***B***.

Proof:

See Theorem 4.6.1.1.

◙

4.8.2 Perfect information is achievable

The following example proves that perfect information is achievable in the set of integrations that satisfy the assumed conditions in this section.

Example 4.8.2.1: Consider two ISs, ***A*** (4.34) and ***B*** (4.35):

(4.34)

Signal /State	y^A_1	y^A_2	y^A_3
s_1	1	0	0
s_2	1	0	0
s_3	0	1	0
s_4	0.1	0.1	0.8

(4.35)

Signal /State	y^B_1	y^B_2	y^B_3
s_1	1	0	0
s_2	0	1	0
s_3	0	1	0
s_4	0.1	0.1	0.8

A and ***B*** are not GMI-related (applying a logic similar to the logic illustrated for ISs (4.21)-(4.23) will yield the desired proof). In addition, it can be inferred from $\boldsymbol{G}^{AB}$=***A@B*** (4.36 below) that ***A*** and ***B*** are not orthogonal. Therefore, the assumptions of the analysis in this section are met.

(4.36)

	y^{AB}_1	y^{AB}_2	y^{AB}_3	y^{AB}_4	y^{AB}_5	y^{AB}_6	y^{AB}_7	y^{AB}_8	y^{AB}_9
s_1	1	0	0	0	0	0	0	0	0
s_2	0	1	0	0	0	0	0	0	0
s_3	0	0	0	0	1	0	0	0	0
s_4	0	0	0.1	0.1	0	0	0	0.1	0.7

G^{AB} is a perfect IS (Lemma 4.3.1.1). ISs ***A*** and ***B*** complement each other's capacity to distinguish between states s_1, s_2, and s_3, and the strong conditional dependence given s_4 has additional positive influence, such that the outcome is perfect information.

◘

4.8.3 No monotonicity

The following example proves that the relationship between the informativeness of the ISs and the informativeness of their integration is not monotonic in the set of integrations that satisfy the assumed conditions in this section.

Example 4.8.3.1: Consider ***A*** (4.34) and ***C*** (4.37):

(4.37)

Signal /State	y^C_1	y^C_2	y^C_3
s_1	1	0	0
s_2	0	1	0
s_3	0	1	0
s_4	0.05	0.05	0.9

C (4.37) is GMI than ***B*** (4.35), in particular, ***M*** such that ***CM=B*** is the following:

(4.38)

Signal /Signal	y^B_1	y^B_2	y^B_3
y^C_1	1	0	0
y^C_2	0	1	0
y^C_3	0.05556	0.05556	0.88888

Suppose that ***G^{AC}*** =***A@C*** is the following:

(4.39)

	y^{AC}_1	y^{AC}_2	y^{AC}_3	y^{AC}_4	y^{AC}_5	y^{AC}_6	y^{AC}_7	y^{AC}_8	y^{AC}_9
s_1	1	0	0	0	0	0	0	0	0
s_2	0	1	0	0	0	0	0	0	0
s_3	0	0	0	0	1	0	0	0	0
s_4	0.004	0.006	0.09	0.004	0.006	0.09	0.04	0.04	0.72

It can be observed based on ***A***, ***C***, and $\boldsymbol{G}^{AC}$ that ***A*** and ***C*** are not orthogonal. It can also be proved that ***A*** and ***C*** are not GMI-related. Therefore, the conditions analyzed in this section are met.

$\boldsymbol{G}^{AB}$ (4.36) is a perfect IS, but $\boldsymbol{G}^{AC}$ (4.39) is not. Therefore, while ***C*** is GMI than ***B***, $\boldsymbol{G}^{AB}$ is GSMI than $\boldsymbol{G}^{AC}$. This example proves that there is no monotonicity under the assumed conditions.

◘

4.8.4 Summary of results: Sources are not state-conditionally independent and cannot be ordered by their quality

Section 4.8 analyzed the informativeness and value of integration when the ISs are not orthogonal, and none has higher informativeness than the other.

The results reflect the combined effect of two conditions, each discussed in turn earlier—non-GMI-related ISs, and state-conditional dependence. The diagram below summarizes the theoretical findings in this section.

	ISs are orthogonal	ISs are *not* orthogonal
One IS is more informative (or ISs equivalent)		
None is more informative		- INCREASE IS GUARANTEED - NO MONOTONICITY - PERFECT INFORMATION IS ACHIEVABLE

4.9 PRELIMINARY THOUGHTS ON THE IMPLICATIONS TO ENSEMBLE LEARNING ALGORITHMS

4.9.1 How does ensemble learning fit in the model of integration of this research?

One way to conceptualize ensemble learning algorithms in terms of information integration as defined by this research, is to view the product as a sequence of *two* information integration mechanisms. The first integration is carried by the single classifiers, i.e., the members of the ensemble. The input of each classifier consists of the values of a chosen set of features, and the output is a class identifier. An IS model of this scenario interprets classes as states; a Cartesian product of the value sets of the chosen features maps to the signal set (finite sets assumed); the integration IS matrix describes the class-conditional joint distributions of the features values. The second integration is carried through majority voting or another group decision making mechanism. It produces a common class identifier based on input consisting of the class identifiers suggested by the individual classifiers. An IS model of the second integration interprets classes, again, as states, and the set of signals is a one-to-one mapping on a Cartesian product of the classes for each of the classifiers.

The IS model of information integration refers to an ideal case, i.e., to the maximal informativeness possible through integration. In practice, a classifier may often offer lower informativeness than the maximum possible based on the selected features, and the peculiarities of the integration mechanism affect the outcome as well.

4.9.2 Integration strategies applied by boosting, error correcting output coding, and bagging

Bagging, boosting, and error correcting output coding, are ensemble learning techniques that use different approaches for increasing the informativeness of classifications.

- The boosting strategy seems to agree with the proposed theory of integration of non-orthogonal ISs.
- The error-correcting output coding strategy appears to agree with the theory of integration of non-GMI-related ISs.
- The bagging approach may form statistically independent classifiers and integrates them; it does not appear to translate to any specific quadrant within the framework of this research.

Boosting (based on Schapire, 2002): Given an instance, boosting aims to control the number of classifiers that identify it correctly, and especially, keep this number high. If the number of correct classifiers is controlled such that, for example, it satisfies majority voting requirements, then the integration mechanism will be accurate.

Ideally, boosting would produce integration that is similar in its fundamental nature to the one depicted in Section 4.7.3, distinguished, in particular, by strong state-conditional dependence between the sources. The conditional probability of many signal combinations would be zero, and the signal combinations that have positive conditional probability would provide perfect information because they have zero probability given the other states.

The boosting algorithm influences the joint conditional probabilities of classifications by constructing the classifiers sequentially and applying a special weighting scheme as part of this process. Specifically, the construction of each classifier is preceded by a step in which the training set that serves in the construction is weighted. The weight of an instance depends on the number of earlier classifiers that identified it correctly. This way, the weight on "hard" instances keeps increasing as the classifier construction process progresses. Since the construction of a classifier is guided by a criterion of minimization of the sum of weighted-errors, the boosting algorithm constructs classifiers with emphasis on the accuracy of higher-weighted instances. Consequently, boosting discourages repeated incorrect classifications.

Error-correcting output coding (based on Dietterich, 2000a): Error-correcting output coding aims to construct complementary classifiers such that each has high informativeness within some proper subsets of the states and no information at all beyond them, and integrate those classifiers. This idea was illustrated in Section 4.6.3 (the examples in Section 4.2.1 underlie the same idea). Before the development of each classifier, the class set is partitioned randomly into two subsets and the subsets are re-labeled (0,1). The training set that serves the construction of a classifier shows the transformed class set. As a result, each classifier focuses on accurate distinction between two random subsets as described. The integration mechanism determines the actual class as the one that is designated by the highest number of classifications.

Bagging (based on Dietterich, 2000a): Bagging evidently aims to create statistically independent classifiers. Diversity of this kind is achieved through random sampling of the training set—the training examples for building a classifier are chosen randomly, with replacements, from the original training set. Due to the sensitivity of many classifier-building algorithms to variations in the training set, such variations can result in substantial differences between classifiers.

Statistical independence is not necessarily equivalent to conditional independence. Bagging does not aim at state-conditional independence per se. It also does not systematically aim at any specific relationship between classifiers in terms of informativeness; different classifiers may correspond either to GMI-related ISs or to non-GMI-related ISs.

4.9.3 Preliminary notes about empirical findings on the success of bagging and boosting

Repeated studies show that boosting often yields higher classification accuracy than bagging when the training set is error-free, although bagging is clearly superior to boosting when the training set is noisy. Here is one possible explanation to the latter observation.

An increasing amount of random noise can modify the conditional distributions associated with the single features towards fixed conditional distributions across the different classes, unconditional independence, and state-conditional independence (e.g., Barabash, 1965). Hence, the ISs that model the different features will be (approximately) orthogonal, such that monotonicity will apply—the input information sources with highest individual accuracy will produce the integration with the highest informativeness.

Bagging applies a strategy that suits the given conditions. First, it constructs the individual classifiers based on random sampling from the training set, such that domain properties are maintained by the sample. A "greedy" search by the single classifiers is appropriate due to the monotonicity property implied by the domain. Subsequently, assuming that the single classifiers are more or less independent and state-conditionally independent, integration using a majority vote mechanism is guaranteed to have a strictly higher maximal informativeness relative to the informativeness of the single classifiers, and, again, maintain a monotonic relationship with the former. (It was proved that majority vote is a suitable integration mechanism under the assumed conditions.)

Boosting, in contrast, seeks to generate conditional dependence among the members of the ensemble that would yield high accuracy of the integration. In the presence of substantial random noise, however, the high weight put on instances with random noise may in fact have a negative effect on the predictive accuracy of the individual classifiers.

In addition, the integration mechanism uses a weighted majority vote that may further magnify any negative influence on accuracy.

The proposed explanation on the effectiveness of boosting versus bagging, under conditions of noise, points to a major issue: the relationship between the characteristics of a domain and the strategy of the ensemble learning algorithm. Part of this issue is the relationship, and combined effect, of the ensemble approach and the approach chosen by individual classifiers.

4.9.4 The link between diversity and classification accuracy

A few studies of ensemble learning methods try to link various measures of "diversity" to classification accuracy (e.g., Hansen and Salamon, 1990; Kuncheva et al. 2003; Kuncheva and Whitaker, 2003; Dietterich 2000b; Ali and Pazzani, 1996). Researchers feel that the improved classification accuracy of ensemble learning algorithms is caused by their "diversity", i.e., disagreement or difference. Higher diversity should thus correspond to higher classification accuracy. However, converse to such beliefs, the results of empirical investigations that applied various measures of diversity and accuracy and examined the relationship between them have been inconclusive.

The findings of this study indicate that current interpretations of the notion of diversity cannot provide a general predictor for variations in classification accuracy. Especially, an attempt to link diversity with statistical independence seems bound to fail. For example, if statistical independence were a good predictor of classification accuracy, it should probably characterize integrations that provide perfect information. However, the findings show that integrations that provide perfect information are often produced by sources that are strongly conditionally dependent, and although such sources may be unconditionally independent, they are not necessarily unconditionally independent. Another example, if statistical independence were a good predictor of classification accuracy, it might be expected that when integrations are not strictly more informative than the sources, there will be complete dependence between the sources. However, current results in regard to the characteristics of integrations that do not have strictly higher informativeness show that the input information sources are not necessarily completely unconditionally dependent.

This study highlights a new interpretation of diversity which links it to differences in domains of informative strength. Extreme instances of such diversity, explained as sources having expertise in complementary areas, were identified under two basic forms, and shown to produce perfect information. Perhaps such findings may help re-shape the approach to diversity.

4.10 CONCLUSIONS

Ahituv and Ronens' findings (1988) mark the beginnings of a theory of information integration using information economics theory.

This research formalized the notion of information integration, and categorized integration scenarios employing two tests: Whether or not the input information sources can be ranked in terms of their informativeness for the problem at hand, and (2) whether or not the input information sources are state-conditionally independent. Four categories were created in this way. Subsequently, three formal criteria on the informativenes and economic value of integration were explored in each of the four categories. These three criteria were termed “increase in informativeness is guaranteed” (/not guaranteed), “monotonicity” (/no monotonicity), and “perfect information is achievable” (/not achievable).

When the information sources are state-conditionally independent there is a monotonic relationship between the informativeness of the sources and the informativeness of their integration. As the informativeness of the sources grows, so does the informativeness of their integration. This outcome fits a source selection strategy that may assist when there is a limit on the number of sources that can take part in the integration: integrate the most informative sources.

Yet, state-conditional dependence between information sources is a powerful factor that typically has to be taken into account together with their individual informativeness. When the state-conditional dependence is allowed to vary, higher informativeness of the input information sources may be accompanied by lower informativeness of their integration—there is no monotonicity.

Sources that are not ranked are often such that each is more informative than the other within some subsets of their value range, i.e., they have complementary informative strengths. This research finds that diversity of this kind is very useful to information integration. The informativeness of the integration is always strictly higher under these conditions than the informativeness of the input sources. Optimal instances of such diversity, portrayed as sources that have expertise in complementary domains, produce perfect information. Each source has perfect information within subsets of the value range, and the sources have complementary structures, i.e., less than perfect information in one matches perfect information in the other.

These conclusions reveal the potential value of a strategy that targets sources that have such complementary strengths. Even when an information source is not highly informative as a whole, "slices" (subsets) of its overall value range may still be highly informative. When different information sources have complementary "slices" of this kind, their integration can always generate positive gain in informativeness, and under the best scenario, the output will reach perfect information quality and value.

Sources that are ranked by their informativeness and are state-conditionally independent are less beneficial for integration then state-conditionally independent sources that are not ranked. Perfect information is not achievable. Furthermore, every integration under these conditions has a boundary on its informativeness that is strictly inferior to perfect information. Finally, the least effective instances are valueless, i.e., the integration is not strictly more informative than all the input sources. These instances include sources that have perfect information within subsets of the value range. Evidently, sources in this category are particularly sensitive to the choice of sources with which they are integrated.

The integration of sources that are ranked by their informativeness can be valueless also when the sources are not state-conditionally independent. The precise characteristics of this undesired scenario have been identified here and may provide help in its prior identification.

On the other hand, strong state-conditional dependence is a necessary condition for the production of perfect information when the sources are not state-conditionally independent. Perfect information appears to be generated under these conditions, as much as under state-conditional independence, by highly diverse source, i.e., such that have complementary expertise. Yet the formal characteristics of diverse sources are substantially different from the comparable situation under state-conditional independence.

Finally, the outcome of this research indicates a possible approach to integration that transforms the integration problem (i.e., decreases or increases the resolution of the model) such that state-conditional independence translates to state-conditional dependence, or vice versa.

The definitions of orthogonality and higher informativeness make strong assumptions on the relationships between the ISs. This way, for example, two ISs in which signals are orthogonal within a proper subset of the states are considered non-orthogonal. However, the current theory can be applied to proper subsets. For example, the understanding that was obtained with respect to orthogonal ISs can be applied to a subset of the states within which the signals are orthogonal. Similarly, the understanding that was obtained with respect to GMI-related ISs can be applied when sub-matrices of the ISs, associated with a proper subset of the states, can be ranked, even though the ISs are not GMI-related.

The investigation also assumes that two ISs are being integrated. Extension to $n \geq 2$ should not pose much difficulty. Example 4.6.3.1 gives a clue to the way in which the theory can be generalized to $n \geq 2$ ISs.

5. SOFTWARE ECONOMICS: DOES IT FOLLOW THE INFORMATION GOODS MODEL?

"At current levels of growth in the tech-support field, early in the new century every person on the planet should be a tech-support specialist."

- Andrew Grove[7]

5.1 INTRODUCTION

Software is typically equated with code in economic analyses in information and software economics. The reality, however, is that buyers complement the code with various services. In many cases software is sold as a package that consists of code and one or more `free` or fee-based services, sometimes including hardware. Traditional services include technical support and training. Other examples involve installation planning, system integration, performance tuning, capacity planning, disaster planning, network planning, network integration, application integration, strategic planning, security consulting, and facility management. Application service providers (ASP's) take an extreme approach to the demand for services, offering application usage as a service, hence sparing the customer from the challenge of code ownership.

Services increasingly form a major share of information technology (IT) markets. IT services comprised more than 40% of worldwide IT spending in 2002, and roughly half

[7] This saying was attributed to Andy Grove by Gary Chapman, in an article that appeared in *Los Angeles Times* on November 27, 2000.

of IT spending in most highly developed countries. A recent commentary by Michael Cusumano (CACM, March 2003) points to the economic importance of services for software producers. Software producers and vendors that are typically perceived as product-oriented are actually heavily invested in services and are earning increasing revenues from selling services. Services are sold either separately, or through "maintenance" packages or "solutions" that bundle code with services and/or hardware.

The high complementarity of code and services can have substantial implications for companies' costs, pricing decisions, differentiation strategies, market entry barriers, and so on, as well as for the competitive conditions in the industry as a whole. However, existing theory including both software economics and the popular information goods model (Shapiro and Varian, 1998) that is often applied to software, equates software with code and ignores the emerging market reality.

We believe that despite the fact that current theory provides useful insights, it fails to explain various phenomena. The conception of software as a bundle or package rather than just code is critical in enriching the understanding the current software market. It may be helpful not only for approaching situations in which software is actually sold as a package, but also in studies that wish to account for the high complementarity of code, service, and hardware.

For example, the information goods model implies that software has high, sunken, first-copy cost, and small, nearly zero, per-copy cost of production. Therefore, pricing based on the marginal cost of production is bound to end in losses; pricing must center on the demand side, i.e., must be based on the buyer's value only. However, data from various sources in the industry suggest that per-package cost can be high enough to affect the price, and in the solutions market-segment this cost is frequently a major determinant of the price. Profit margins on solutions and service packages have been significantly lower than expected, partly due to cost management difficulties. Further, heavy price reductions may go together with some form of unbundling as a way of recovering non-code costs.

The bundling view of the economics of software may be perceived to be related to the recently popular term “total cost of ownership” (TCO). However, the concept of TCO is from the perspective of the buyer/owner who is seeking to reduce the whole life cost of the product; TCO typically takes the supplier strategy as given. The focus of this paper, in contrast, is on the supplier’s strategy.

5.2 BUNDLING PATTERNS: DEMAND AND SUPPLY

Packages that combine code and service vary widely in their content. Despite the popularity of bundles in the marketplace, data regarding the actual bundles provided are not discussed in the software economics literature. Therefore, it is useful to provide a brief overview of the major trends in software bundling.

Maintenance plans. A growing number of software producers require their customers to purchase the code together with a maintenance plan that covers the first year. A recent survey (Association of Support Professionals, 2002) estimates that 34% of all software producers have this requirement. A maintenance plan itself is a package that generally provides telephone and/or e-mail technical support, access to restricted Web-based support materials, and a guarantee of provision of any major code upgrades available in the time-period of the plan. Some vendors require customers to purchase a maintenance plan as long as the product is in use.

The spreading popularity of mandatory maintenance among vendors is primarily explained by the stable flow of revenue that maintenance offers, unlike the uncertainty involved in selling software licenses—especially when the economy is down. Oracle's recent financial report clearly reflects this advantage. Its first fiscal quarter for 2003 was quite poor in terms of new license sales, which dropped almost 25% relative to the same quarter last year. However, Oracle's overall decrease in revenue was only a little over l0%, largely due to the fact that new license sales comprised only 28% of its total revenue and revenue from maintenance grew during that quarter. Another claimed advantage of maintenance plans for the vendor is reduced support cost. Keeping all the customers under one version eliminates the need to track and support different versions and thereby lowers support cost.

Customers, on their part, demand maintenance plans even when they are optional. A survey (Association of Support Professionals, 2002) found that 95% (median) of the customers purchase maintenance plans for the first year, and 86% (median) of those renew the contract in the following year. This trend may be explained, at least in part, by a common attitude among CIOs that (Slater, 1999): "if something breaks, CIOs want someone they can call to fix it—or to strangle if whatever's wrong can't be fixed in an acceptable time frame."[8]

---

[8] Originally this attitude was suggested to explain the lack of popularity of Linux and OSS among business organizations at a time when Linux was not backed by strong technical support.

Free technical support. In the PC market, especially among software for individual or small business use, code is sometimes bundled with some guarantee of free personal support. Historically, it was common for PC software producers to bundle code with unlimited telephone support free of charge. Unlike the increasing popularity of maintenance plans, guarantees of free personal support are gradually diminishing, despite strong resistance of the buyers in this market segment who are reluctant to pay for support. Notably, the drastic software price cuts in the early 90's accelerated this shift.

Today, software publishers that target individual users or small companies provide a mix of free and fee-based support, in which free telephone support, if offered, has a limited term. Software producers have shifted free support to the Internet, based on the assumption that the marginal cost of an incident should decline substantially as the interaction of users with human support technicians decreases. Some online alternatives decrease the interaction through automation; others, like user forums, utilize free human information sources outside the organization.

Solutions. Defined as “A combination of products and services, focused on a particular business or technology problem, which drives measurable business value” (Dover, 2002), solutions are recent trend in packaging that emphasizes bottom line results for the customer, as well as the application of a variety of capabilities for that purpose. It implies a “customer-centric” rather than a “product-centric” approach.

The rising demand for solutions is related to the growing complexity of information technology that has left many organizations without the technical skills that are needed in order to utilize new products. The implementation of E-commerce, ERP (Enterprise Resource Planning), Y2K projects, CRM (Customer Relationship Management), and other large systems posed a challenge that was beyond what many companies could handle. As a result the demand for “traditional” services like technical support and training, as well as “professional” services such as those mentioned in the introduction, grew. There has also been an increasing demand for complete turnkey solutions to specified business needs.

Vendors have seen the demand for entire solutions as an opportunity for product differentiation, allowing them to maintain and increase their revenues, increase control of their market, and/or improve their falling profit margins. Subsequently, there has been a significant shift among both product and service organizations to become solutions providers. A solution may be delivered fully by one organization, or involve a close cooperation between two or even more organizations. It can be standardized (fixed content, price, delivery dates), or tailored to the customer's special needs. Sometimes it includes a hardware component, other times a hardware component is an option, or not offered at all. The shift to services and solutions sometimes involves un-bundling or re-bundling of offerings. Services that in the past were bundled with products and given away free of charge to “keep a customer happy”, are now enhanced, augmented, and un-bundled or re-bundled with products—on a fee basis.

5.3 COST OF SERVICE

The information goods model suggests that software is costly to produce but cheap to reproduce—it generates high sunk, first-copy cost and very low, nearly zero, per-copy cost for additional copies. The cost of a package that combines code with service is, however, affected by the properties of the cost of service. This cost has a large variable component caused by labor requirements.

Labor costs are the major contributor to the overall cost of technical support, which has been estimated to be around 8% of the revenues of software publishers (see Table 5.1) (Soft Letter, 1997). This estimate varies widely, up to 20% of revenues, depending on application type, software price, operation size, and business model in use i.e., fee-based support or free support.

A report by Cusumano and Selby on Microsoft's telephone support (Cusumano and Selby, 1995) indicates to five major determinants of the cost of labor in technical support. These factors are consistent with the notion of the variable nature of this cost.

- The average length of a call
- Number of calls
- Variations in the number of calls over time
- Employee pay
- The nature of the product

TABLE 5.3: Support cost ratios

	Support employee % (Median)	Support cost % (Median)
By company size:		
$100+ million	11%	7.0%
$10+ million	15%	6.0%
$5-$10 million	17.5%	15.0%
under $5 million	20%	20.0%
By support organization size:		
30+ employees	15.5%	8.0%
10-29 employees	10.0%	6.7%
1-9 employees	10.0%	10.0%
By percentage of fee-based calls:		
75%+ fee-based	15.0%	10.0%
75%+ free	15.0%	7.0%
By price of best-selling product:		
$10,000+	15.0%	10.0%
$1000-$9,995	15.0%	12.5%
Under $1000	15.0%	5.0%
By application category:		
System software	15.0%	15.0%
Vertical/industry- or job-specific	19.0%	12.0%
Programming tool/language	12.0%	8.0%
Accounting/finance/tax	32.5.0%	20.0%
Communications/networking	10.0%	5.0%
General business	12.5%	7.0%
Utility	20%	10.0%
Consumer/education	17%	3.0%
Median (all)[9]	**15%**	**8.1%**

SOURCES: Soft-letter/Association of Support Professionals "Technical Support Cost Ratios" 1997, 2000.

Despite many efforts in the software industry to lower labor requirements, data show that related services are still labor-intensive. Table 5.3 shows survey findings (Soft Letter,

[9] This median corresponds to most of the data in this table. The only exception is the data set under the classification by application.

1997) on the percentage of PC software developers' workforce dedicated to technical support. The overall median is 15%, with notable variations across application category and company size (economies of scale).

In addition to technical support, companies offer a variety of other services. The percentage of a software producer's workforce dedicated to services as a whole can actually be substantially higher than what the data in Table 5.3 suggest. As the examples below indicate, services may take a higher portion of a producer's workforce than code development (R&D).

- Novell's annual report for 2001 informs about 2869 workers in services (44%) compared to 1293 in R&D (20%).
- Peoplesoft's annual report for 2001 tells about 3732 workers in services (44%), and 2214 in R&D (26%).
- Microsoft's documents do not reveal how many of its employees are consultants and/or technical support staff. However, a former head of Microsoft's worldwide services reported on 14,000 consultants and support technicians at the end of 2001. Around 9000 of those were support technicians. Microsoft's financial records for 2001 and 2002 indicate that over 28% of its workforce are on the service side compared to nearly 42% R&D staff (20,800).
- Oracle's financial report for the year that ended in May 31, 2002, reports on 8,859 R&D workers - 21% of its total workforce. Oracle, which is more service-oriented

than Microsoft, does not specify the number of support technicians and consultants. Sales and service taken together, however, Oracle employed 27,059 (64%) workers.

The cost of labor among professional service jobs is much higher than among technical support jobs, due to more demanding skill constraints. Salaries that are twice as high are very common. Professional service projects are often also much more time consuming—weeks, months, or even years long.

In some cases the combined costs of all services can be the highest item among a software producer's cost categories. For example:

- Oracle's cost of service in 2001 was $2.8 billion—higher than its cost of sales and marketing ($2.69 billion), product development ($1.14), and administration ($0.46).
- Peoplesoft's cost of service was $.70 billion in 2001, compared to $.36 billion in cost of product development and license fees, $.52 billion—sales and marketing, and $.16 billion—administration costs.

Many product-oriented organizations entering the service arena suffer from inefficiency and elevated costs due to lack of experience. Service operation, service sales and marketing, and management, are very different from the corresponding product functions. Furthermore, solutions require special approaches to marketing, sales, and delivery (Dover, 2002).

It seems clear that the available data support the notion that software packages have a significant variable cost, similar to a service cost model and unlike the information goods cost model.

5.4 BUNDLE PRICE AND PRICING

Given the evidence that there are significant variable costs associated with the cost of a software bundle, one would expect that there would be many cases that illustrate the influence of the variable cost of the service on the package price and/or content. We present the following cases for illustration purposes:

<u>Price cuts and maintenance fees.</u> Maintenance charges have typically followed a simple model—the price of maintenance is set as a percentage (usually 15%-25%) of the purchase price of the code. Heavy discounts, however, pose a special risk under this approach. Unlike the marginal cost of code, which diminishes quickly with the number of copies, the marginal cost of support does not diminish and is also uncertain at the time of purchase. Over time this pricing model is being adjusted in a way that avoids the downside risk by setting the maintenance fee as a percentage of the list price of the code, rather than its purchase price (Association of Support Professionals, 2002).

<u>Price cuts and free technical support.</u> In the early 90's the PC software market experienced dramatic price cuts. By the end of 1993 PC software vendors had great difficulty keeping technical support free: “We used to be able to sell software packages for $200. The price of support was basically built into it…[but] when that software package gets down to $50, $30, or less, there was no way for us to be able to pay for the cost too.”(Wilson, 2001). When Microsoft introduced new “priority”, and “premium” fee-based plans in the fall of 1983, many other software companies immediately increased paid support options and limited free support. The vendors' approach to the cost

problem was to explicitly unbundle the code/support package rather than maintain higher bundle prices and run the risk of incurring large variable support costs.

TABLE 5.4: Code/technical support bundles and their prices

Company	Type of free telephone support	Application type	Price range
Adobe	One free incident.	Productivity	$49 - $122
	Free support for a period of 90 days.	Productivity	$199 and up
Connectix	No free telephone support	Utility Communications	$29-$99
	One free incident	Utility Communications	$99-$249
Corel	One free call on installation and configuration issues.	Productivity	$18 - $129[10]
	Free support for a period of 30 days, limited to installation and configuration issues.	Productivity	$79-$489
Interplay	Free unlimited telephone support	Games	$5-$50
Intuit	Free unlimited telephone support	Tax	$30-$100
	Free telephone support on installation issues.	Accounting; Tax	$60- $500
Macromedia	No free telephone support	Productivity	$99
	Free support for a period of 90 days.	Productivity	$299 and up
Microsoft	Free unlimited telephone-support	Games; Educational	Below $100
		Windows (operating system)	
		Office (productivity)	
		Other productivity applications	
Ontrack Data International	No free telephone support	Utility	$20-$5070
Scansoft	One free incident on installation issues	Utility	$59-$995
Symantec	No free telephone support	Utility	$30-$180
Take-Two Interactive Software	Free unlimited telephone-support	Games	$10-$50

[10] These prices were collected on www.amazon.com, and www.computer4sure.com.

Table 5.4 shows data that were collected in July 2002 from the websites of eleven software producers in the consumer and home market segment.[11] It describes the free telephone support that these vendors bundle with different software types, and matching price ranges. Price clearly offers only partial explanation to the variation in telephone support offerings. Nonetheless, in four of the companies price is positively related to the extent of telephone support (Adobe, Connectix, Corel, and Macromedia). Inquiries to game producers and Intuit confirmed the assumption that low support cost was associated with high product usability and quality and an important factor in the decision to offer unlimited, free support (see also the Consumer/Education application category in Table 5.3).

Solutions. Industry experts recommend that services and solution providers set prices based on the value to the customer, yet they also note this is often not the case. Prices are determined on an estimated cost basis (even though many companies do not understand their cost structure very well) or in relationship to comparable offerings by competitors.

Service labor costs were forecasted to be around 60% of revenues at present and gross profit margins were expected to be similar to those of traditional services (around 35%), or more (Mikelson, 2001; Saleh, 2002). Actual profit margins are estimated to be around 15%-20% (Mikelson, 2001). This situation is partly explained by the general lack of experience in this domain, as mentioned earlier.

---

[11] The companies were selected from a list of top 100 desktop application producers produced by *softletter* (2000), and the data correspond to the offerings of these producers for the home segment.

In contrast to the predictions of the "software as code" paradigm, the marginal cost of a solution package, especially labor cost, is an important factor in pricing. Often it is, in fact, a major determinant of price.

5.5 Conclusions

We believe that a conception of software as a package or bundle can be valuable in approaching a broad set of economic phenomena, and can balance the current theory that interprets software mainly as code. It may be helpful not only for approaching situations in which software is actually sold as a package, but also in studies that wish to account for the high complementarity of code, service, and hardware.

Contrary to the information goods model, our data point to conditions in which the marginal cost of software is high enough to affect, or even, largely determine, price. Furthermore, profit margins of solutions have been lower than expected, partly due to cost management difficulties in relation to lack of experience with this model. A recurring approach of vendors, when the price of a bundle goes down significantly which is demonstrated with bundles that include an element of technical support, has been to do away with the bundle one way or another.

A potentially important factor that influences pricing that was not discussed in this paper is the tradeoff in production between code and service: the demand for services is, in general, dependent on decisions made during code development. Together with our argument, this implies that a software firm's investments in quality can increase the firm's bundle pricing options for the future.

6. CONCLUSIONS AND FUTURE DIRECTIONS

The widely used database technology and recent developments in networking and Web technologies are encouraging diversity in the utilization of existing data. Today, data are routinely pooled from multiple systems and physical locations, and integrated for various analytical and decision-making purposes. From a managerial perspective, however, there are growing concerns in regard to the quality of the output information, and the economic justification of costly investments in such technologies.

Much of this dissertation addresses related issues: the effect of integration on the quality and value of information, and the effect of data quality improvements on the quality and value of information. In addition, the final part of the dissertation considers a question concerning software economics. Chapter 6 offers an overview of the results, contribution, and suggestions for future research.

6.1 RESULTS AND CONTRIBUTION

6.1.1 Assessing the informativeness and value of information integration

Chapter 4 presents the outcome of a formal investigation of the effect of information integration on the quality and value of information. Information economics theory provides the conceptual and analytical foundation for this analysis.

The framework in use classifies information integration situations based on two information quality characteristics—*informativeness* and *dependence*. More precisely, the study categorizes integration scenarios employing two tests that take for granted the state set that is targeted by the information:

1. Whether or not the input information sources can be ranked in terms of their informativeness for the problem at hand
2. Whether or not the input information sources are statistically state-conditionally independent.

Four categories are created in this way.

The inquiry focuses on three criteria on the informativenes and economic value of integration:

4. Whether or not an increase in informativeness is guaranteed through integration.
5. Whether or not higher informativeness of the individual ISs guarantees higher informativeness of their integration ("monotonicity").
6. Whether or not perfect information is achievable through integration.

Consistent with the IS theory, the formal definition of integration and other concepts designate optimal outcome.

The analysis assumes that two information sources are being integrated. It can be extended to *n* information sources. Table 6.5 summarizes the high-level findings.

	Sources are state-conditionally independent	Source are *not* state-conditionally independent
One source is more informative (or sources equivalent)	-INCREASE IS NOT GUARANTEED - MONOTONICITY - PERFECT INFORMATION IS NOT ACHIEVABLE	- INCREASE IS NOT GUARANTEED - NO MONOTONICITY - PERFECT INFORMATION IS ACHIEVABLE
None is more informative	-INCREASE IS GUARANTEED - MONOTONICITY - PERFECT INFORMATION IS ACHIEVABLE	- INCREASE IS GUARANTEED - NO MONOTONICITY - PERFECT INFORMATION IS ACHIEVABLE

TABLE 6.5: Summary of high-level results on information integration

When the information sources are state-conditionally independent there is a monotonic relationship between the informativeness of the sources and the informativeness of their integration. As the informativeness of the sources grows, so does the informativeness of their integration. This outcome points to a source selection strategy that may assist when there is a limit on the number of sources that can take part in the integration: integrate the most informative sources. Such strategy has the potential to yield the highest informativeness and economic value under the assumed conditions.

Yet, state-conditional dependence is a powerful factor that typically has to be taken into account together with the individual informativeness of the sources. When the state-conditional dependence is allowed to vary, higher informativeness of the input information sources may be accompanied by lower informativeness of their integration—there is no monotonicity.

Sources that are not ranked by their informativeness may be such that each is more informative than the other within some subsets of their value range, i.e., they have complementary informative strengths. This research finds that diversity of this kind is very useful to information integration. The informativeness of the integration is always strictly higher under these conditions than the informativeness of the input sources. Optimal instances of such diversity, portrayed as sources that have expertise in complementary domains, produce perfect information. Each source has perfect information within subsets of the value range, and the sources have complementary structures, i.e., less then perfect information in one matches perfect information in the other.

These conclusions emphasize the potential value of a strategy that targets sources that have such complementary strenghts. Even when an information source is not highly informative as a whole, "slices" (subsets) of its overall value range may still be highly informative. When different information sources have complementary "slices" of this kind, their integration can always generate positive gain in informativeness, and under the best scenario, the output will reach perfect information quality and value.

Sources that are state-conditionally independent and one is more informative than the other may be common among information sources that are plagued with random noise. They are less beneficial for integration then state-conditionally independent sources that cannot be ranked by their informativeness. Perfect information is not achievable.

Furthermore, each integration has a boundary on its informativeness that is strictly inferior to perfect information. Finally, the least effective instances are valueless. These involve sources that, again, have perfect information within subsets of the states. Evidently, sources in this category are particularly sensitive to the choice of sources with which they are integrated.

The integration of sources that are ranked by their informativeness can be valueless also when the sources are not state-conditionally independent. The precise characteristics of this undesired scenario have been identified here too, and may provide help in its prior identification.

On the other hand, strong state-conditional dependence is a necessary condition for the production of perfect information when the sources are not state-conditionally independent. Perfect information appears to be generated under these conditions, as much as under state-conditional independence, by highly diverse source, i.e., such that have complementary expertise. Yet the formal characteristics of diverse sources are substantially different from the comparable situation under state-conditional independence.

The outcome of this research indicates to a possible approach to integration that transforms the integration problem (i.e., decreases or increases the resolution of the model) such that state-conditional independence translates to state-conditional dependence, or vice versa.

This research contributes to current understanding in the following ways:

1. It offers a theoretical framework to assist analysis and meta-analysis of information integration problems.
2. It develops and applies a set of general analytical tools based on information economics theory. Such tools enrich the set of tools for use with information economics theory.
3. Some of the theoretical findings appear to be completely new in the context of related literature. In particular, past research has ignored the possibility that two sources cannot be ranked as none is more informative than the other.

6.1.2 Does higher data accuracy produce higher forecast accuracy?

Chapter 3 and Chapter 4 address the question of whether higher accuracy of an information system's input ensures higher accuracy of its output. Specifically, Chapter 3 and Chapter 4 complement each other in an investigation of conditions that imply such "monotonicity" or, alternatively, "non-monotonicity".

IS theory provides the conceptual and analytical foundation for a major part of this analysis. Consistent with IS theory, the analysis assesses variations in the maximal accuracy, informativeness, and expected utility of the output. Another part of the analysis concentrates on a type of non-classical measurement errors in the context of a simple linear regression model.

Chapter 3 investigates the research question under the assumption that the information system uses a single input source for generating the output information. The information system is viewed as a function that maps the values of a one-dimensional variable to the values of a one-dimensional variable.

Chapter 4 addresses a similar problem under the assumption that the information system produces information through synthesis of multiple input sources. The information system is associated with a function that maps the values of a multi-dimensional variable to the values of a one-dimensional variable.

The study identifies a contextual factor that can affect monotonicity. This factor is *dependence between errors*. A sufficient condition for monotonicity is state-conditional independence between errors.

Chapter 3 distinguishes one type of such dependence between errors, which is particularly meaningful when the relationship between the input and output variables is stochastic. This kind of dependence relationship is illustrated in a simple regression model by the dependence between the *measurement error* in the independent variable and the *unexplained error*.

When the relationship between an information system's input and output variables is deterministic, the errors of interest in Chapter 3 are state-conditionally independent. Therefore, this type of error does not obstruct monotonicity in deterministic settings. However, an interpretation of the results of Chapter 4 from a data quality perspective directs to another type of dependence between errors that is meaningful in deterministic setting, and can also affect the outcome in stochastic settings. The dependence that

Chapter 4 recognizes might be illustrated by the relationship between measurement errors in different independent variables in a linear regression model.

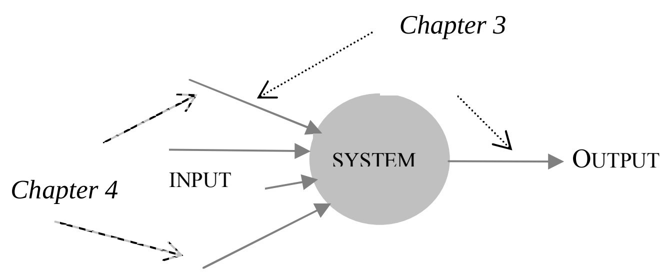


FIGURE 6.3: Dependence relationships that affect monotonicity

Chapter 3 and Chapter 4 offer various examples that prove that when the sufficient conditions for monotonicity are not satisfied, higher input accuracy can go together with lower output accuracy, i.e., there is no monotonicity.

This research contributes to current understanding in the following ways:

1. The findings of this inquiry establish a generalization of current data quality theory in MIS. Instead of the deterministic orientation of earlier work, this research accounts for both stochastic and deterministic relationships between the phenomena designated by the input and output of an information system.

2. The study develops and applies a set of general analytical tools based on information economics theory. Such tools enrich the set of tools for use with information economics theory.

6.1.3 Software economics: Does it follow the information goods model

Despite the major share that services take of the software industry, the economic implications of the demand and supply of services are not well understood at this stage. The chapter on software economics aims at this gap.

Code and service are highly complementary. Accordingly, the study examines a bundling view of the economics of software—a conception of software as a bundle or package that consists of code, services, and possibly also hardware, rather than just code. This view is explored in a specific domain—the basic propositions of the information goods model. Contrary to the information goods model, the data that was collected from various sources in the industry suggest that per-package cost can be high enough to affect the price, and in the solutions market-segment this cost is frequently a major determinant of the price. Profit margins on solutions and service packages have been significantly lower than expected, partly due to cost management difficulties. Further, heavy price reductions may go together with some form of unbundling as a way of recovering non-code costs.

This research contributes to current understanding in the following ways:

1. A conception of software as a bundle may enrich the theory that addresses the supplier's side, which currently interprets software mainly as code and ignores the market reality. The findings provide preliminary justification to this conception.

2. The findings challenge a popular model in software economics thinking.

6.2 FUTURE RESEARCH DIRECTIONS

6.2.1 Assessing the informativeness and value of information integration

The theoretical framework developed in this study may have a variety of applications. It can serve as a supporting framework for a methodology or a technonology that implements integration; it can also be useful in the review of research findings in relevant domains, and in directing practical decision-making situations where the issue is integration.

On the other hand, the understanding of the informativeness and value of information integration is only partial, such that further theoretical investigation can be valuable. One direction is driven by the uneven distinction that the chosen framework makes between orthogonal ISs, and non-orthogonal ISs. In contrast to the concept of orthogonal ISs that refers to a unique dependence pattern, non-orthogonal ISs group together an infinite number of dependence patterns. A more elaborate investigation of the latter domain can use evaluation criteria similar to the criteria employed by this research (e.g., monotonicity), or suggest new evaluation criteria.

The literature review (Chapter 2) reveals a variety of research areas where similar issues are of interest. Three potential domains will be addressed next: classification/prediction model-building algorithms, in particular ensemble learning algorithms; assessment and resolution of inconsistent overlapping data in loose integration of databases; and economic analysis of information systems integration projects—especially information sharing in the supply chain.

<u>Learning algorithms.</u> Learning algorithms, such as decision-trees, or neural networks, are in essence integration mechanisms that, typically, take a set of features as inputs and produce class labels (in classification tasks) or real numbers (in prediction tasks). The current framework may be useful in explaining some of the empirical results pertaining to learning algorithms, and possibly also direct some future research in this domain.

For example, in regard to a recurring effort to predict the accuracy of ensemble learning based on "diversity". Measures of diversity portray the differences in classifications in an environment with multiple classifiers. However, attempts to predict accuracy based on diversity have not been successful so far. The findings of this research indicate that current interpretations of the notion of diversity cannot serve as broad predictors of variations in classification accuracy. Instead, this study highlights another interpretation of diversity, which associates it with differences in domains of informative strength.

The results indicate to the importance of the right match between the learning strategy and the properties of the underlying domain. This way, for example, empirical results that show that "boosting" type of ensemble learning has poorer performance than "bagging" in domains with substantial random noise may be explained from the standpoint of the interaction between the learning strategy, and the characteristics of the domain. In essence, the relatively simple approach of bagging that applies "greedy" algorithms without manipulating them may have better capacity to exploit the monotonicity property of the domain than the control mechanism that boosting employs over the individual classifiers, which may not be so useful when variations are random.

Addressing data quality in a loose integration of databases. Some research that aims to resolve inconsistencies when data from multiple sources overlap through explicit meta-data and reasoning about quality was done in the Multiplex project (e.g., Anokhin and Motro, 1999; Motro and Rakov, 1998, Rakov, 1997).

Few studies partition the database, through the use of views, into areas with homogenous quality. This study offers theoretical justification to that approach via its results on ISs that cannot be ranked. The results on ISs that cannot be ranked also advise that the integration mechanism identify highly complementary "slices" and select them for integration. This strategy has not been implemented yet, apparently.

The problem of inconsistencies is approached under the implicit assumption of monotonicity. That is, the quality of individual inputs is taken as an indicator to the quality of the product of integration. The findings imply that better quality may be achieved through a more discerning approach to quality assessment and inconsistency resolution that recognizes the implications of state-conditional dependencies.

Economic analysis of information systems integration projects. A significant number of recent inquiries have addressed the question of the value of information sharing in the supply chain, using formal analyses and other methodologies. This literature focuses on demand uncertainties, so the objective of information sharing is, in particular, higher accuracy of future demand forecasts.

Current research is apparently based on an implicit assumption that the quality of information about demand is uniform regardless of the actual state (e.g., the forecasting error is normally distributed with zero mean and variance σ^2, as in Aviv, 2001). Research also appears to assume that the outcome of a forecasting error in terms of payoff is the same, regardless of the actual state.

In these circumstances, the new framework may encourage consideration of variations in information quality and their implications. Such variation may be relevant for example, for considering information about new trends in demand separate from known trends (e.g., new patterns may be relatively hard to predict). The framework may also direct a finer scrutiny of the origins of higher information quality and value, e.g., to what extent is it due to "complementary expertise" in different state subsets?

6.2.2 Does higher data accuracy produce higher forecast accuracy

Viewed as a theory of data quality and value, the current findings invite further exploration along two general avenues: theoretical, and methodological/technical.

From the perspective of theory, an investigation of environments in which error patterns that affect monotonicity might be common can be useful. There is a need to develop classification of such error patterns in practical settings, assess their predominance, assess their economic effect, and so on. Research on these questions may be carried in various ways, e.g., using theoretical analyses, simulations, and surveys.

Empirical findings on non-classical measurement errors in the social sciences indicate that similar error patterns are caused by the tendency of people to give false information, believing that they can gain from lying. Intuition suggests that people demonstrate such tendencies in organizational settings as well. Consequently, game-theoretic models that examine different information strategies within various organizational settings may help direct the search for environments in which the interesting error-patterns can be found. Theoretical analyses may utilize other models to analyze the possible economic effects of such error patterns.

Scenario #1 offers some insight that may be valuable for a methodological approach to causes of unwanted dependence between errors. The scenario links dependence between errors with the existence of a hidden factor, i.e., a factor that is not accounted for. Although the model that marketing analysts produced may reflect the true relationship between the units in current use and future demand variables ($d^*_2 = 0.5d^*_1 + \varepsilon$), such model overlooks a relevant factor, namely, the managerial strength of the organization. An approach for surfacing such hidden variables and handle them one way or the other might be useful.

Empirical studies and practical applications of this theory will require an ability to assess error patterns. Assessment of this type may be based on a sample of accurate data. For example, in the illustrative scenario (Scenario #1 in Chapter 1) such sample might be produced by an extended version of the survey described there. An extension would ask for data on the number of equipment units in current use by the customer. This way, the equipment producer would be able to compare the data obtained through the survey (assumably accurate) to customer database records, in order to estimate the error.

See also previous comments with respect to research directions on information integration.

6.2.3 Software Economics: Does it indeed follow the Information Goods Model?

Future research can explore the effect of the high complementarity of code and service on software pricing in further depth. For example, how is the price of code related to the price of service when they are not sold as a package? Under what conditions is it profitable (not profitable) to bundle the code with a given service that the producer offers? How does reliance on outside service provider(s), e.g., consultant companies, affect software price? More broadly, how does a partnership with consulting firms affect the market strategies of software producers? Investigation along such lines may utilize analytical models, as well as empirical methodologies.

A potentially important influences on pricing which was not discussed in this paper is the tradeoff in production between code and service: the demand for services is, in general, dependent on decisions made during code development. Together with the argument of the study, this tradeoff implies that a software firm's investments in quality can increase the firm's pricing options for the future. Investments in higher software quality, as well as other software design decisions may be analyzed from this perspective.

Finally, the high complementarity of code and service appears to affect the software industry in other ways. For example, it may establish higher entry barriers in some cases, as may have been the case with the operating system Linux. The issue of higher entry barriers may deserve a separate examination.

APPENDIX: A COMPLETE PROOF OF THEOREM 4.5.1.1

Theorem 4.5.1.1. ***A***, ***B***, denote ISs defined on SxY_A, SxY_B, respectively, such that ***A*** and ***B*** are orthogonal. An increase in informativeness is guaranteed for ***A*** and ***B*** if:

2. There exists a two row sub-matrix of ***B***, ***B'***, in which the rows correspond to $S' \subseteq S$, such that ***B'*** is neither a null IS nor a perfect IS
3. ***A'***, a two row sub-matrix of ***A*** whose rows correspond to $S' \subseteq S$, is neither a null IS nor a perfect IS.

Proof:

It will be shown that ***A*** is not GMI than ***G***=***A***@***B***. Therefore, due to the symmetry of ***A*** and ***B***, ***B*** is not GMI than ***G*** either. Since ***G*** is GMI than ***A*** and ***B*** (Theorem 4.4.1), it follows, by definition, that ***G*** is GSMI than ***A*** and ***B***. Therefore, by definition, an increase in informativeness is guaranteed for ***A*** and ***B***.

Part II

This case assumes ISs ***A***, ***B***, that are square matrices, i.e., det ***A***≠0, det ***B***≠0. Let r denote the number of columns and rows in ***A*** and ***B***. Elements of ***A***, ***B***, will be denoted such that ***A***=$[a_{i,j}]$, ***B***=$[b_{i,j}]$, i=1,.., r, j=1,.., r. The elements of the integration of ***A*** and ***B***, ***G***=***A***@***B***, will be denoted such that ***G***=$[a_{i,j} \cap b_{i,k}]$, i=1,.., r, j=1,.., r, k=1,.., r.

To keep notation short I compromise on consistency. The notation $a_{1,1}$, $b_{1,1}$,.. for values of ***A*** and ***B*** designates conditional probabilities. The notation of the values of ***G***, e.g., $a_{1,1} \cap b_{1,1}$ designate the respective joint conditional probabilities.

Suppose that, contrary to the theorem's conclusion, ***A*** is GMI than ***G***, such that there exists a Markov matrix ***M***, ***AM***=***G***. Due to the fact that ***AM***=***G*** corresponds to r^2 equation sets all having ***A*** as the coefficient matrix, and det ***A***≠0, the existence of a single solution to each equation set is guaranteed (Simon and Blume, 1994, Fact 7.10). A global analysis of the r^2 solutions follows, in order to establish the conditions under which the solutions form a Markov matrix, as required.

The number of rows of ***M*** is equal to the number of columns of ***A***, i.e., r , and the number of columns of ***M***, r^2 , is equal to the number of columns of ***G***. An element of ***M***, $m_{i,j}$, is calculated based on Cramer's rule in the following manner: column i of ***A*** is replaced by column j of ***G*** , and the determinant of the subsequent matrix, denoted here $\boldsymbol{A}^{\boldsymbol{G}}_{i,j}$, is then divided by the determinant of ***A***.

This way, for example:

$$m_{11}=\det \boldsymbol{A}^{G}{}_{1,1}\div \det \boldsymbol{A}=\det \begin{vmatrix} a_{11}\cap b_{11} & a_{12} & . & . & . & a_{1r} \\ a_{21}\cap b_{21} & . & . & . & . & . \\ & & & . & . & . \\ & & & . & . & . \\ . & . & . & & & \\ a_{r-1,1}\cap b_{r-1,1} & . & . & & & \\ a_{r\,1}\cap b_{r1} & a_{r2} & & & & a_{rr} \end{vmatrix} \div \det \boldsymbol{A}$$

Let the (i,j) minor of $\boldsymbol{A}$ be denoted by $\boldsymbol{A}(i,j)$. Elements of the first row of $\boldsymbol{M}$ are then given by:

$m_{1,1}=[a_{1,1}\cap b_{1,1} \det \boldsymbol{A}^{G}{}_{1,1}(1,1) - a_{2,1}\cap b_{2,1} \det \boldsymbol{A}^{G}{}_{1,1}(2,1) + a_{3,1}\cap b_{3,1} \det \boldsymbol{A}^{G}{}_{1,1}(3,1) -..]\div\det \boldsymbol{A}$
$m_{1,2}=[a_{1,1}\cap b_{1,2} \det \boldsymbol{A}^{G}{}_{1,2}(1,1) - a_{2,1}\cap b_{2,2} \det \boldsymbol{A}^{G}{}_{1,2}(2,1) + a_{3,1}\cap b_{3,2} \det \boldsymbol{A}^{G}{}_{1,2}(3,1) -..]\div\det \boldsymbol{A}$
$m_{1,3}=[a_{1,1}\cap b_{1,3} \det \boldsymbol{A}^{G}{}_{1,3}(1,1) - a_{2,1}\cap b_{2,3} \det \boldsymbol{A}^{G}{}_{1,3}(2,1) + a_{3,1}\cap b_{3,3} \det \boldsymbol{A}^{G}{}_{1,3}(3,1) -..]\div\det \boldsymbol{A}$
etc.

Suppose that the first r elements of the first row of $\boldsymbol{M}$ are aggregated. Since $\boldsymbol{A}^{G}{}_{1,1}(1,1)= \boldsymbol{A}^{G}{}_{1,2}(1,1)= \boldsymbol{A}^{G}{}_{1,3}(1,1)=..=\boldsymbol{A}(1,1)$; $\boldsymbol{A}^{G}{}_{1,1}(2,1)= \boldsymbol{A}^{G}{}_{1,2}(2,1)= \boldsymbol{A}^{G}{}_{1,3}(2,1)=..= \boldsymbol{A}(2,1)$;.., and - $\sum_{i=1}^{r} a_{1,1}\cap b_{1,i}=a_{1,1}$, $\sum_{i=1}^{r} a_{2,1}\cap b_{2,i}=a_{2,1}$,.. A summation of the first r elements yields:

$\sum_{i=1}^{r} m_{1,r}= [a_{1,1} \det \boldsymbol{A}(1,1) - a_{2,1} \det \boldsymbol{A}(2,1) + a_{3,1} \det \boldsymbol{A}(3,1) -..\pm a_{r,1} \det \boldsymbol{A}(r,1)] \div \det \boldsymbol{A} =$
$\det \boldsymbol{A} \div \det \boldsymbol{A} = 1.$

Similarly, the second group of r elements of the first row of $\boldsymbol{M}$ consists of:

$m_{1,r+1}=[a_{1,2}\cap b_{1,1}\det \boldsymbol{A}^{G}{}_{1,r+1}(1,1) - a_{2,2}\cap b_{2,1} \det \boldsymbol{A}^{G}{}_{1,r+1}(2,1) + a_{3,2}\cap b_{3,1} \det \boldsymbol{A}^{G}{}_{1,r+1}(3,1)-..]\div \det \boldsymbol{A}$
$m_{1,r+2}=[a_{1,2}\cap b_{1,2} \det \boldsymbol{A}^{G}{}_{1,r+2}(1,1) - a_{2,2}\cap b_{2,2} \det \boldsymbol{A}^{G}{}_{1,r+2}(2,1) + a_{3,2}\cap b_{3,2} \det \boldsymbol{A}^{G}{}_{1,r+2}(3,1)-..]\div \det \boldsymbol{A}$
$m_{1,r+3}=[a_{1,2}\cap b_{1,3} \det \boldsymbol{A}^{G}{}_{1,r+3}(1,1) - a_{2,2}\cap b_{2,3} \det \boldsymbol{A}^{G}{}_{1,r+3}(2,1) + a_{3,2}\cap b_{3,3} \det \boldsymbol{A}^{G}{}_{1,r+3}(3,1)-..]\div \det \boldsymbol{A}$
...

The aggregate -

$\sum_{i=1}^{r} m_{1,r+i} = [a_{1,2} \det \boldsymbol{A}(1,1)- a_{2,2} \det \boldsymbol{A}(2,1) + a_{3,2} \det \boldsymbol{A}(3,1) -..\pm a_{r,2} \det \boldsymbol{A}(r,1)] \div \det \boldsymbol{A}$
is however equal to zero in this case since the expression in the numerator corresponds to a matrix with linearly dependent columns (a matrix that is the same as $\boldsymbol{A}$ except that the first and second columns are identical and equal to the first column of $\boldsymbol{A}$). As a result,

since $\boldsymbol{M}$ is a Markov matrix and therefore its values are non-negative, each of the component elements is equal zero too.

$m_{1,r+1} = [a_{1,2} \cap b_{1,1} \det \boldsymbol{A}(1,1) - a_{2,2} \cap b_{2,1} \det \boldsymbol{A}(2,1) + a_{3,2} \cap b_{3,1} \det \boldsymbol{A}(3,1) - ..] \div \det \boldsymbol{A} = 0$
$m_{1,r+2} = [a_{1,2} \cap b_{1,2} \det \boldsymbol{A}(1,1) - a_{2,2} \cap b_{2,2} \det \boldsymbol{A}(2,1) + a_{3,2} \cap b_{3,2} \det \boldsymbol{A}(3,1) - ..] \div \det \boldsymbol{A} = 0$
$m_{1,r+3} = [a_{1,2} \cap b_{1,3} \det \boldsymbol{A}(1,1) - a_{2,2} \cap b_{2,3} \det \boldsymbol{A}(2,1) + a_{3,2} \cap b_{3,3} \det \boldsymbol{A}(3,1) - ..] \div \det \boldsymbol{A} = 0$
…

and since $\det \boldsymbol{A} \neq 0$,

$m_{1,r+1} = a_{1,2} \cap b_{1,1} \det \boldsymbol{A}(1,1) - a_{2,2} \cap b_{2,1} \det \boldsymbol{A}(2,1) + a_{3,2} \cap b_{3,1} \det \boldsymbol{A}(3,1) - .. = 0$
$m_{1,r+2} = a_{1,2} \cap b_{1,2} \det \boldsymbol{A}(1,1) - a_{2,2} \cap b_{2,2} \det \boldsymbol{A}(2,1) + a_{3,2} \cap b_{3,2} \det \boldsymbol{A}(3,1) - .. = 0$
$m_{1,r+3} = a_{1,2} \cap b_{1,3} \det \boldsymbol{A}(1,1) - a_{2,2} \cap b_{2,3} \det \boldsymbol{A}(2,1) + a_{3,2} \cap b_{3,3} \det \boldsymbol{A}(3,1) - .. = 0$
…

When $\boldsymbol{A}$ and $\boldsymbol{B}$ are orthogonal, the above equations have a special form:

$m_{1,r+1} = a_{1,2} b_{1,1} \det \boldsymbol{A}(1,1) - a_{2,2} b_{2,1} \det \boldsymbol{A}(2,1) + a_{3,2} b_{3,1} \det \boldsymbol{A}(3,1) - .. = 0$
$m_{1,r+2} = a_{1,2} b_{1,2} \det \boldsymbol{A}(1,1) - a_{2,2} b_{2,2} \det \boldsymbol{A}(2,1) + a_{3,2} b_{3,2} \det \boldsymbol{A}(3,1) - .. = 0$
$m_{1,r+3} = a_{1,2} b_{1,3} \det \boldsymbol{A}(1,1) - a_{2,2} b_{2,3} \det \boldsymbol{A}(2,1) + a_{3,2} b_{3,3} \det \boldsymbol{A}(3,1) - .. = 0$
…

This particular form represents, however, a homogeneous system of r linear equations, with r unknowns, namely, $a_{1,2} \det \boldsymbol{A}(1,1)$, $-a_{2,2} \det \boldsymbol{A}(2,1)$, $a_{3,2} \det \boldsymbol{A}(3,1)$,.. The coefficients form the matrix $\boldsymbol{B}^T$. Therefore, since $\det \boldsymbol{B} \neq 0$, the system has only one solution by which $a_{1,2} \det \boldsymbol{A}(1,1) = a_{2,2} \det \boldsymbol{A}(2,1) = a_{3,2} \det \boldsymbol{A}(3,1) = .. = 0$.

Due to the fact that $\det \boldsymbol{A} \neq 0$, it is not possible that $\det \boldsymbol{A}(1,1) = \det \boldsymbol{A}(2,1) = \det \boldsymbol{A}(3,1) = .. = \det \boldsymbol{A}(r,1) = 0$.

Suppose therefore, that one of these minors, which, without loss of generality, can be assumed to be $\boldsymbol{A}(1,1)$ is such that $\det \boldsymbol{A}(1,1) \neq 0$. Therefore $a_{1,2}=0$.

Next, take the third group of r elements in the first row of $\boldsymbol{M}$:

$m_{1,2r+1} = [a_{1,3} \cap b_{1,1} \det \boldsymbol{A}^G_{1,2r+1}(1,1) - a_{2,3} \cap b_{2,1} \det \boldsymbol{A}^G_{1,2r+1}(2,1) + a_{3,3} \cap b_{3,1} \det \boldsymbol{A}^G_{1,2r+1}(3,1) - ..] \div \det \boldsymbol{A}$
$m_{1,2r+2} = [a_{1,3} \cap b_{1,2} \det \boldsymbol{A}^G_{1,2r+2}(1,1) - a_{2,3} \cap b_{2,2} \det \boldsymbol{A}^G_{1,2r+2}(2,1) + a_{3,3} \cap b_{3,2} \det \boldsymbol{A}^G_{1,2r+2}(3,1) - ..] \div \det \boldsymbol{A}$
$m_{1,2r+3} = [a_{1,3} \cap b_{1,3} \det \boldsymbol{A}^G_{1,2r+3}(1,1) - a_{2,3} \cap b_{2,3} \det \boldsymbol{A}^G_{1,2r+3}(2,1) + a_{3,3} \cap b_{3,3} \det \boldsymbol{A}^G_{1,2r+3}(3,1) - ..] \div \det \boldsymbol{A}$
etc.

Due to the fact that the aggregate of these values is equal to zero too, they, likewise, must be zero. Therefore:

$m_{1,2r+1}= a_{1,3}\cap b_{1,1}\det \boldsymbol{A}^{G}_{1,2r+1}(1,1) - a_{2,3}\cap b_{2,1} \det\boldsymbol{A}^{G}_{1,2r+1}(2,1) + a_{3,3}\cap b_{3,1} \det\boldsymbol{A}^{G}_{1,2r+1}(3,1) -..= 0$
$m_{1,2r+2}= a_{1,3}\cap b_{1,2}\det \boldsymbol{A}^{G}_{1,2r+2}(1,1) - a_{2,3}\cap b_{2,2} \det\boldsymbol{A}^{G}_{1,2r+2}(2,1) + a_{3,3}\cap b_{3,2} \det\boldsymbol{A}^{G}_{1,2r+2}(3,1) -..= 0$
$m_{1,2r+3}= a_{1,3}\cap b_{1,3}\det \boldsymbol{A}^{G}_{1,2r+3}(1,1) - a_{2,3}\cap b_{2,3} \det\boldsymbol{A}^{G}_{1,2r+3}(2,1)+ a_{3,3}\cap b_{3,3} \det\boldsymbol{A}^{G}_{1,2r+3}(3,1) -..= 0$
…

Which, when $\boldsymbol{A}$ and $\boldsymbol{B}$ are orthogonal, is the same as:

$m_{1,2r+1} = a_{1,3}b_{1,1} \det \boldsymbol{A}^{G}_{1,2r+1}(1,1) - a_{2,3}b_{2,1} \det \boldsymbol{A}^{G}_{1,2r+1}(2,1) + a_{3,3}b_{3,1} \det \boldsymbol{A}^{G}_{1,2r+1}(3,1) -..= 0$
$m_{1,2r+2} = a_{1,3}b_{1,2} \det \boldsymbol{A}^{G}_{1,2r+2}(1,1) - a_{2,3}b_{2,2} \det \boldsymbol{A}^{G}_{1,2r+2}(2,1) + a_{3,3}b_{3,2} \det \boldsymbol{A}^{G}_{1,2r+2}(3,1) -..= 0$
$m_{1,2r+3} = a_{1,3}b_{1,3} \det \boldsymbol{A}^{G}_{1,2r+3}(1,1) - a_{2,3}b_{2,3} \det \boldsymbol{A}^{G}_{1,2r+3}(2,1)+ a_{3,3}b_{3,3} \det \boldsymbol{A}^{G}_{1,2r+3}(3,1) -..= 0$
…

The above set is a homogeneous system of r linear equations, with r unknowns, i.e., $a_{1,3} \det \boldsymbol{A}(1,1)$, $-a_{2,3} \det \boldsymbol{A}(2,1)$, $a_{3,3} \det \boldsymbol{A}(3,1)$, .. Again, since the coefficients of this system form the matrix $\boldsymbol{B}^{T}$, and $\det \boldsymbol{B}\neq 0$, the system has only one solution, by which $a_{1,3} \det \boldsymbol{A}(1,1) = a_{2,3} \det \boldsymbol{A}(2,1) = a_{3,3} \det \boldsymbol{A}(3,1) = .. = 0$. But, as before, due to the fact that $\det \boldsymbol{A}\neq 0$, it is not possible that $\det \boldsymbol{A}(1,1) = \det \boldsymbol{A}(2,1) = \det \boldsymbol{A}(3,1) = .. = \det \boldsymbol{A}(r,1) = 0$.

Moreover, carrying on with the earlier assumption that $\det \boldsymbol{A}(1,1)\neq 0$, it must be that $a_{1,3}=0$. It follows that, altogether, $a_{1,2}=0$, and $a_{1,3}=0$.

The same logic can be repeated, to produce the following result: All the values of one row of $\boldsymbol{A}$ except for one value are equal zero. For example, under the assumption made here, $a_{1,1}\neq 0$, $a_{1,2}=0$, $a_{1,3}=0$, $a_{1,4}=0$.. $a_{1,r}=0$.

Next, the second row of $\boldsymbol{M}$ undergoes the same analysis. A summation of the first r elements yields zero, which implies, when $\boldsymbol{A}$, $\boldsymbol{B}$, are orthogonal, the following homogeneous system of r linear equations with r unknowns:

$m_{2,1} = -a_{1,1}b_{1,1} \det \boldsymbol{A}(1,2) + a_{2,1}b_{2,1} \det \boldsymbol{A}(2,2) - a_{3,1}b_{3,1} \det \boldsymbol{A}(3,2) +.. = 0$
$m_{2,2} = -a_{1,1}b_{1,2} \det \boldsymbol{A}(1,2) + a_{2,1}b_{2,2} \det \boldsymbol{A}(2,2) - a_{3,1}b_{3,2} \det \boldsymbol{A}(3,2) +.. = 0$
$m_{2,3} = -a_{1,1}b_{1,3} \det \boldsymbol{A}(1,2) + a_{2,1}b_{2,3} \det \boldsymbol{A}(2,2) - a_{3,1}b_{3,3} \det \boldsymbol{A}(3,2) +.. = 0$
…

Once again, the system has exactly one solution, in which all the unknowns are zero.
$a_{1,1} \det \boldsymbol{A}(1,2) = a_{2,1} \det \boldsymbol{A}(2,2) = a_{3,1} \det \boldsymbol{A}(3,2) = .. = 0$.

As explained, at least one of the determinants must be non-zero. In addition, to be consistent with the initial assumption that $\det \boldsymbol{A}(1,1)\neq 0$, $a_{1,1}\neq 0$, $a_{1,2}=0$, $a_{1,3}=0$, $a_{1,4}=0$,..

$a_{1,r}$=0, it cannot be that det ***A***(1,2)≠0, because this implies $a_{1,1}$= 0. Therefore, say, det ***A***(2,2) = 0, such that $a_{2,1}$=0.

A corresponding analysis can show next that all the values of the second row of ***A*** are zero, except for $a_{2,1}$.

And so on.

This process can be repeated *r* times on each row of ***M***, since ***M*** has are r^2 columns. This way all the rows of ***A*** are examined, one by one, with a comparable result. Therefore, as a whole, ***A*** is a Markov matrix, det ***A***≠0, and every column of ***A*** has exactly one value that is not zero. In other words, ***A*** is a perfect IS (Lemma 4.3.1.1). But if ***A*** is a perfect IS then the conditions of this theorem are not satisfied.

Therefore ***A*** is not GMI than ***G***.

c columns	*c* columns	*c* columns	*c* columns	
$m_{1,1}m_{1,2}$..	0 0 0……….0	0 0 0……….0		
0 0 0……….0	$m_{2,r+1}m_{2,r+2}$..	0 0 0……….0		
0 0 0……….0	0 0 0……….0	$m_{3,2r+1}m_{3,2r+2}$.		
.	.	.		
.	.	0 0 0……….0		• • •
.	.	.		
.	.	.		
0 0 0……….0	0 0 0……….0	0 0 0……….0		

FIGURE A.1: The structure imposed on ***M***

Part III:

Case 1: The rank of ***A*** is equal to the number of columns of ***A*** (which implies that the number of columns of ***A*** is less than, or equal to, the number of rows there).

As before, the proof, by contradiction, examines the constraints on any Markov matrix ***M*** such that ***AM***=***G***, ***G***=***A***@***B***, and, consequently, the constraints on ***G***, ***A***, and ***B***.

Let *r* denote the number of rows of ***A***, let *c'* denote the number of columns of ***A*** (*c'*≤*r*), and let *c* denote the number of columns of ***B***. It is assumed that all the columns of ***G*** are linearly dependent on the columns of ***A***.

Under the described conditions there exists a single matrix ***M*** such that ***AM***=***G*** (Simon and Blume, 1994, Fact 7.2 and Fact 7.9) The values of ***M*** can be found by solving *cc'* equation sets, each of them having a coefficient matrix $\tilde{\boldsymbol{A}}$ that is a square sub-matrix of ***A*** that consists of *c'* linearly independent rows of ***A***. The right-hand side values of these equation sets form a sub-matrix derived from the respective rows of $\boldsymbol{G}$. This sub-matrix is denoted next $\hat{\boldsymbol{G}}$. It has c' rows and cc' columns. In other words, $\tilde{\boldsymbol{A}}\boldsymbol{M}=\hat{\boldsymbol{G}}$.

Given $\tilde{\boldsymbol{A}}$ and $\hat{\boldsymbol{G}}$, the previous analysis of the square matrix case applies. Therefore a Markov matrix $\boldsymbol{M}$ such that $\tilde{\boldsymbol{A}}\boldsymbol{M}=\hat{\boldsymbol{G}}$, $\boldsymbol{AM}=\boldsymbol{G}$, has the structure as depicted in Figure A.1. Therefore $\boldsymbol{G}$ has the structure as depicted in Figure 4.2. Due to the special structure imposed on $\boldsymbol{G}$, its jth sequence of c columns is such that the columns are dependent only on the jth column of $\boldsymbol{A}$. Hence, the rank of the sub-matrix formed by these columns is equal 1. In addition, when $\boldsymbol{A}$ and $\boldsymbol{B}$ are orthogonal, as assumed here, the values in $\boldsymbol{G}$ are products of respective values in $\boldsymbol{A}$ and $\boldsymbol{B}$. This way, for example, the 1st sequence of c columns in $\boldsymbol{G}$ is given by:

$$\begin{array}{cccc} a_{11}b_{11} & a_{11}b_{12} & \ldots.. & a_{11}b_{1c} \\ a_{21}b_{21} & a_{21}b_{22} & \ldots.. & a_{21}b_{2c} \\ \ldots.. & \ldots.. & \ldots.. & \ldots\ldots \\ a_{r1}b_{r1} & a_{r1}b_{r2} & \ldots.. & a_{r1}b_{rc} \end{array}$$

If $\boldsymbol{B}$ is a null IS then the rank of the above sub-matrix of $\boldsymbol{G}$ is indeed 1. However, in that case the conditions of this theorem are not satisfied. Therefore, without loss of generality, suppose that the first two rows of $\boldsymbol{B}$ are not identical such that $b_{1,1}\neq b_{2,1}$. Suppose also that $a_{1,1}\neq 0$, $a_{2,1}\neq 0$. If the rank of the above sub-matrix is 1, it must be that $a_{1,1}/a_{2,1}=b_{2,1}/b_{1,1}=b_{2,2}/b_{1,2}=\,..=b_{2,c}/b_{1,c}$. But since $b_{2,1}/b_{1,1}\neq 1$, it follows that $b_{2,2}/b_{1,2}\neq 1$, $b_{2,3}/b_{1,3}\neq 1$, etc. Therefore $\boldsymbol{B}$ is not a Markov matrix, because either the sum of the values in the first row, or the second row, of $\boldsymbol{B}$, is not equal one. But this contradicts the assumptions on $\boldsymbol{B}$. Therefore, $a_{1,1}=0$, or $a_{2,1}=0$.

In a similar way it can be shown that $b_{1,1}\neq b_{2,1}$ implies that $a_{1,2}=0$ or $a_{2,2}=0$, $a_{1,3}=0$ or $a_{2,3}=0$, and so on. Therefore, using Lemma 4.3.1.1, the first two rows of $\boldsymbol{A}$ define a perfect IS in contradiction to the conditions of this theorem.

Therefore $\boldsymbol{A}$ is not GMI than $\boldsymbol{G}$.

<u>Case 2</u>: The rank of $\boldsymbol{A}$ is less than the number of columns of $\boldsymbol{A}$.

As before, the proof, by contradiction, examines the constraints on any Markov matrix $\boldsymbol{M}$ such that $\boldsymbol{AM}$=$\boldsymbol{G}$, $\boldsymbol{G}$=$\boldsymbol{A}$@$\boldsymbol{B}$, and, consequently, the constraints on $\boldsymbol{G}$, $\boldsymbol{A}$, and $\boldsymbol{B}$. Similar to the previous analyses, it will be shown that the columns in the jth sequence of c' columns in $\boldsymbol{G}$ have rank 1. The same logic as in the previous analysis can then be applied to prove that $\boldsymbol{A}$ is not GMI than $\boldsymbol{G}$.

Let r denote the rank of $\boldsymbol{A}$, let c denote the number of columns of $\boldsymbol{A}$, and let c' denote the number of columns of $\boldsymbol{B}$. It is assumed that all the columns of $\boldsymbol{G}$ are linearly dependent on the columns of $\boldsymbol{A}$, therefore one or infinitely many solutions are guaranteed for a matrix $\boldsymbol{M}$ such that $\boldsymbol{AM}$=$\boldsymbol{G}$ (Simon and Blume, 1994, Fact 7.2 and Fact 7.11). Again, the proof examines the constraints on a Markov matrix $\boldsymbol{M}$ such that $\boldsymbol{AM}$=$\boldsymbol{G}$, $\boldsymbol{G}$=$\boldsymbol{A}$@$\boldsymbol{B}$, and, consequently, the constraints on $\boldsymbol{G}$, $\boldsymbol{A}$, and $\boldsymbol{B}$.

$\boldsymbol{M}$ has c rows and cc' columns. The values of $\boldsymbol{M}$ can be found by solving cc' equation sets, each of them having a coefficient matrix $\boldsymbol{A'}$ that is a sub-matrix of $\boldsymbol{A}$ that consists of r linearly independent rows of $\boldsymbol{A}$. To simplify the logic it will be assumed that the first r rows and r columns of $\boldsymbol{A}$ are linearly independent. These assumptions are not significant in any other way (e.g., see Marschak, 1971, Section 4.8).

Like in Case 1 above, $\boldsymbol{A'M}$=$\boldsymbol{G'}$ where $\boldsymbol{G'}$ consists of the first r rows of $\boldsymbol{G}$.

Given the solution values of the last c-r variables in each column of $\boldsymbol{M}$, the other values of $\boldsymbol{M}$ can be derived by solving $\tilde{\boldsymbol{A}}\boldsymbol{M'}$=$\hat{\boldsymbol{G}}$, where $\tilde{\boldsymbol{A}}$ consists of the first r columns of $\boldsymbol{A'}$, $\boldsymbol{M'}$ refers to the first r rows of $\boldsymbol{M}$, and $\hat{\boldsymbol{G}}$ consists of the values of $\boldsymbol{G'}$, adjusted in the following way. The value $a_{1,1}\cap b_{1,1}$ are adjusted to $a_{1,1}\cap b_{1,1}-\sum_{i=r+1}^{c}a_{1,i}m_{i,1}$; the value $a_{2,1}\cap b_{2,1}$ are adjusted to $a_{2,1}\cap b_{2,1}-\sum_{i=r+1}^{c}a_{2,i}m_{i,1}$, and so on.

Assuming these adjustments, the elements of the first row of $\boldsymbol{M'}$ are described by:

$$m_{1,1}=[(a_{1,1}\cap b_{1,1}-\sum_{i=r+1}^{c}a_{1,i}\,m_{i,1})\det\tilde{\boldsymbol{A}}^{G}_{1,1}(1,1)-(a_{2,1}\cap b_{2,1}-\sum_{i=r+1}^{c}a_{2,i}\,m_{i,1})\det\tilde{\boldsymbol{A}}^{G}_{1,1}(2,1)+..]\div\det\tilde{\boldsymbol{A}}$$

$$m_{1,2}=[(a_{1,1}\cap b_{1,2}-\sum_{i=r+1}^{c}a_{1,i}\,m_{i,2})\det\tilde{\boldsymbol{A}}^{G}_{1,2}(1,1)-(a_{2,1}\cap b_{2,2}-\sum_{i=r+1}^{c}a_{2,i}\,m_{i,2})\det\tilde{\boldsymbol{A}}^{G}_{1,2}(2,1)+..]\div\det\tilde{\boldsymbol{A}}$$

$$m_{1,3}=[(a_{1,1}\cap b_{1,3}-\sum_{i=r+1}^{c}a_{1,i}\,m_{i,3})\det\tilde{\boldsymbol{A}}^{G}_{1,3}(1,1)-(a_{2,1}\cap b_{2,3}-\sum_{i=r+1}^{c}a_{2,i}\,m_{i,3})\det\tilde{\boldsymbol{A}}^{G}_{1,3}(2,1)+..]\div\det\tilde{\boldsymbol{A}}$$

...

Since $\tilde{\boldsymbol{A}}^{G}_{1,1}(1,1)$=$\tilde{\boldsymbol{A}}^{G}_{1,2}(1,1)$=$\tilde{\boldsymbol{A}}^{G}_{1,3}(1,1)$=.. =$\tilde{\boldsymbol{A}}(1,1)$; $\tilde{\boldsymbol{A}}^{G}_{1,1}(2,1)$=$\tilde{\boldsymbol{A}}^{G}_{1,2}(2,1)$=$\tilde{\boldsymbol{A}}^{G}_{1,3}(2,1)$=..=$\tilde{\boldsymbol{A}}(2,1)$; .. and due to the fact that $\sum_{j=1}^{c'}b_i=1$, a summation of the first c' elements of the first row of $\boldsymbol{M'}$ will yield the value:

$$\sum_{i=1}^{c'} m_{1,i} = \{[a_{1,1} \det \tilde{A}(1,1) - a_{2,1} \det \tilde{A}(2,1) + a_{3,1} \det \tilde{A}(3,1) -..\pm a_{r,1} \det \tilde{A}(r,1)] \div \det \tilde{A}\} -$$

$$\{[\sum_{i=r+1}^{c} a_{1,i} \sum_{j=1}^{c'} m_{i,j} \det \tilde{A}(1,1) - \sum_{i=r+1}^{c} a_{2,i} \sum_{j=1}^{c'} m_{i,j} \det \tilde{A}(2,1) ..\pm \sum_{i=r+1}^{c} a_{r,i} \sum_{j=1}^{c'} m_{i,j} \det \tilde{A}(r,1)] \div \det \tilde{A}\} =$$

$$\mathbf{1} - \{[\sum_{i=r+1}^{c} a_{1,i} \sum_{j=1}^{c'} m_{i,j} \det \tilde{A}(1,1) - \sum_{i=r+1}^{c} a_{2,i} \sum_{j=1}^{c'} m_{i,j} \det \tilde{A}(2,1) ..\pm \sum_{i=r+1}^{c} a_{r,i} \sum_{j=1}^{c'} m_{i,j} \det \tilde{A}(r,1)] \div \det \tilde{A}\}$$

However, each of the last $c - r$ columns of $\boldsymbol{A}$ is a linear combination of the columns of $\tilde{\boldsymbol{A}}$. Therefore, the aggregate above can be re-expressed in terms of the columns of $\tilde{\boldsymbol{A}}$. Let $\boldsymbol{k}_n^i$ denote the coefficient of column n ($n \le r$) in the linear combination that forms column i ($r< i \le c$) of $\boldsymbol{A}$.

Then, the above aggregate is equal :

$$\mathbf{1} - \{[\sum_{i=r+1}^{c} (\sum_{n=1}^{r} \boldsymbol{k}_n^i a_{1,n}) \sum_{j=1}^{c'} m_{i,j} \det \tilde{A}(1,1) - \sum_{i=r+1}^{c} (\sum_{n=1}^{r} \boldsymbol{k}_n^i a_{2,i}) \sum_{j=1}^{c'} m_{i,j} \det \tilde{A}(2,1) ..$$

$$\pm \sum_{i=r+1}^{c} (\sum_{n=1}^{r} \boldsymbol{k}_n^i a_{r,i}) \sum_{j=1}^{c'} m_{i,j} \det \tilde{A}(r,1)] \div \det \tilde{A}\} =$$

since all the columns of $\tilde{\boldsymbol{A}}$ except for its first column generate a zero determinant when they are positioned in the first column, the expression above is reduced to:

$$= \mathbf{1} - \{[\sum_{i=r+1}^{c} \boldsymbol{k}_1^i a_{1,1} \sum_{j=1}^{c'} m_{i,j} \det \tilde{A}(1,1) - \sum_{i=r+1}^{c} \boldsymbol{k}_1^i a_{2,1} \sum_{j=1}^{c'} m_{i,j} \det \tilde{A}(2,1) ..$$

$$\pm \sum_{i=r+1}^{c} \boldsymbol{k}_1^i a_{r,1} \sum_{j=1}^{c'} m_{i,j} \det \tilde{A}(r,1)] \div \det \tilde{A}\} = \mathbf{1} - \sum_{i=r+1}^{c} \boldsymbol{k}_1^i \sum_{j=1}^{c'} m_{i,j}$$

Similarly, a summation of the following c' elements of the first row of $\boldsymbol{M'}$ yields the value:

$$\sum_{i=1}^{c'} m_{1,c'+i} = \{[a_{1,2} \det \tilde{A}(1,1) - a_{2,2} \det \tilde{A}(2,1) + a_{3,2} \det \tilde{A}(3,1) -..\pm a_{r,2} \det \tilde{A}(r,1)] \div \det \tilde{A}\} -$$

$$\{\sum_{i=r+1}^{c} a_{1,i} \sum_{j=1}^{c'} m_{i,c'+j} \det \tilde{A}(1,1) - \sum_{i=r+1}^{c} a_{2,i} \sum_{j=1}^{c'} m_{i,c'+j} \det \tilde{A}(2,1)..\pm \sum_{i=r+1}^{c} a_{r,i} \sum_{j=1}^{c'} m_{i,c'+j} \det \tilde{A}(r,1)] \div \det \tilde{A}\}$$

$$= \mathbf{0} - \{\sum_{i=r+1}^{c} a_{1,i} \sum_{j=1}^{c'} m_{i,c'+j} \det \tilde{A}(1,1) - \sum_{i=r+1}^{c} a_{2,i} \sum_{j=1}^{c'} m_{i,c'+j} \det \tilde{A}(2,1)..\pm \sum_{i=r+1}^{c} a_{r,i} \sum_{j=1}^{c'} m_{i,c'+j} \det \tilde{A}(r,1)] \div$$

$$\det \tilde{A}\} = \mathbf{0} - \sum_{i=r+1}^{c} \boldsymbol{k}_1^i \sum_{j=1}^{c'} m_{i,c'+j}$$

and so on. The resultant aggregates, except for the last c-r aggregates, have a similar form, i.e., zero minus a weighted sum of the values in the last c-r rows in the respective columns. The last c-r aggregates have a somehow different form, e.g., aggregate r+1 in the first row is

$$\sum_{i=1}^{c'} m_{1,rc'+i} = \boldsymbol{k}_1^{r+1} - \sum_{i=r+1}^{c} \boldsymbol{k}_1^{i} \sum_{j=1}^{c'} m_{i,rc'+j}$$

Turning to the second row of $\boldsymbol{M'}$, the corresponding aggregates are

$$\mathbf{0} - \sum_{i=r+1}^{c} \boldsymbol{k}_2^{i} \sum_{j=1}^{c'} m_{i,j}\ ;\ \mathbf{1} - \sum_{i=r+1}^{c} \boldsymbol{k}_2^{i} \sum_{j=1}^{c'} m_{i,c'+j}\ ;\ \mathbf{0} - \sum_{i=r+1}^{c} \boldsymbol{k}_2^{i} \sum_{j=1}^{c'} m_{i,2c'+j}\ ;\ \ldots \boldsymbol{k}_2^{r+1} - \sum_{i=r+1}^{c} \boldsymbol{k}_2^{i} \sum_{j=1}^{c'} m_{i,rc'+j} \ldots$$

etc. The expressions that define the elements in other rows of $\boldsymbol{M'}$ obey the same pattern as the aggregates in the first two rows.

Next, it will be shown more specifically that the sub-matrix of $\boldsymbol{M'}$ formed by its first rc' columns has the familiar diagonal structure depicted in Figure A.1. Furthrmore, it will be proved that deviations from the diagonal structure in $\boldsymbol{M}$ outside the specified sub-matrix are limited, if any, such that the structure of $\boldsymbol{G}$ is the same as in Figure 4.1.

In order for $\boldsymbol{M}$ to be a Markov matrix, each of the above aggregates must be non-negative and less than or equal to one; the sum of the values in each row has to be one. The next step is therefore an analysis of the different possible aggregates with respect to their compliance with the Markov constraints. The analysis will be based on a classification of the linear dependencies that each of the last c-r columns of $\boldsymbol{A}$ generates.

- Negatively linearly dependent columns:

Suppose that a negative dependence on a certain column is unique, i.e., there are no other negative dependencies on that column. For example, suppose that column r+1 is the only column negatively dependent on the second column of $\boldsymbol{A}$. Therefore, $\boldsymbol{k}_2^{i}$ is negative, but no other $\boldsymbol{k}_2^{r+1}$ is negative.

Consider $\boldsymbol{k}_2^{r+1} - \sum_{i=r+1}^{c} \boldsymbol{k}_2^{i} \sum_{j=1}^{c'} m_{i,rc'+j}$. It can be easily seen that the only values of $\boldsymbol{M}$ that will satisfy the Markov matrix requirements on this expression are such that $\sum_{j=1}^{c'} m_{i,rc'+j}$ is zero for every i where $\boldsymbol{k}_2^{i}$ is positive, except for i=r+1 for which $\sum_{j=1}^{c'} m_{i,rc'+j} = 1$. The value of $\boldsymbol{k}_2^{r+1} - \sum_{i=r+1}^{c} \boldsymbol{k}_2^{i} \sum_{j=1}^{c'} m_{i,rc'+j}$ will then be zero.

Therefore, block r+1 of $\boldsymbol{M}$ will be all zeros (i.e., c' zeros) in its second row and in any of the last c-r rows that are multiplied by a positive dependency coefficient, except for row r+1. Row r+1 will have non-negative values, such that block r+1 will amount to one, and the rest of the elements in row r+1 will be zero, since, in a Markov matrix, any row must have a total of one.

But the elements of row r+1 appear in other aggregates too. Therefore, any "zero" type aggregate, i.e., such that begins with the digit zero, will have a zero value corresponding to $\boldsymbol{k}_1^{r+1}$, $\boldsymbol{k}_2^{r+1}$, $\boldsymbol{k}_3^{r+1}$ . . .

For example, in $\mathbf{0} - \sum_{i=r+1}^{c} \boldsymbol{k}_2^i \sum_{j=1}^{c'} m_{i,j}$, the sum $\sum_{j=1}^{c'} m_{r+1,j}$ will be zero (corresponding to $\boldsymbol{k}_2^{r+1}$), and therefore this (negative) coefficient will have no influence on the zero type aggregate. An aggregate type "one", e.g., $\mathbf{1} - \sum_{i=r+1}^{c} \boldsymbol{k}_2^i \sum_{j=1}^{c'} m_{i,c'+j}$ will be affected in a similar fashion.

It can be similarly shown that, in general, negative dependencies have no influence on any "zero" as well as "one" type of aggregate. It can be shown that the first r blocks of $\boldsymbol{M}$ are affected in a similar way when there is more than one negative dependence on a certain column of $\boldsymbol{A}$.

- Positive dependence

Every column among the last c-r columns of $\boldsymbol{A}$ will be positively dependent on one or more of the first r columns of $\boldsymbol{A}$. Therefore, suppose, without loss of generality, that column r+1 is dependent on the first column of $\boldsymbol{A}$, such that $\boldsymbol{k}_1^{r+1} > 0$.

Since negative dependencies have no influence on any "zero" type of aggregate, it follows that the influence of any positive dependence on a "zero" type of aggregate in which it is involved must be zero too. Otherwise the aggregate will be negative, violating the requirements on a Markov matrix. Therefore, the value of a "zero" type aggregate is, invariably, zero. Therefore, due to non-negativity considerations again, there are zeros in r-1 of the first r blocks for any row of $\boldsymbol{M'}$. Moreover, since "zero" type aggregates are all off the diagonal of $\boldsymbol{M'}$, the diagonal structure as in Figure A.1 is maintained in the first r blocks of $\boldsymbol{M'}$.

For $\mathbf{0} - \sum_{i=r+1}^{c} \boldsymbol{k}_1^i \sum_{j=1}^{c'} m_{i,c'+j}$ to be non-negative it must be that the aggregates that appear in the expression and are multiplied by non-zero coefficients are all zeros. Since, by assumption $\boldsymbol{k}_1^{r+1} > 0$, it follows that $\sum_{j=1}^{c'} m_{r+1,c'+j}$ is zero. Therefore, due to non-negativity constraints, the c' elements in row r+1 of the second block of $\boldsymbol{M}$ are all zeros.

In fact, $\boldsymbol{k}_1^{r+1} > 0$ produces r-1 "zero" aggregates in row r+1 of $\boldsymbol{M}$, similar to what was shown for the second block of c' elements there. Consequently, except for the first block of c' values which corresponds to a "one" type of aggregate, the values in r-1 blocks out of the first r blocks in that row are equal zero.

Suppose that column r+1 of $\boldsymbol{A}$ is linearly dependent on another column among the first r columns of $\boldsymbol{A}$, e.g., $\boldsymbol{k}_2^{r+1} > 0$. Therefore, since the aggregate of the first c' elements in the second row of $\boldsymbol{M}$ is a "zero" type aggregate, the first c' elements in row r+1 of $\boldsymbol{M}$ must be zero too. Therefore, all the values in the first r blocks of row r+1 are zero. (The aggregate of the first c' elements is of type "zero" in any of the last r-1 rows of $\boldsymbol{M'}$, so the choice of the second row does not limit the generality of this conclusion.)

Since every column among the last c-r columns of $\boldsymbol{A}$ is positively dependent on at least one of the first r columns of $\boldsymbol{A}$, the structure of the first r blocks of $\boldsymbol{M}$ emerges as follows. The first r rows are block diagonal, where each block consists of one row, c' columns; values outside these blocks are zero. For any column i among the last c-r columns of $\boldsymbol{A}$ that is linearly dependent on two or more of the first r columns of $\boldsymbol{A}$, the first r blocks in row i of $\boldsymbol{M}$ are all zero. For any column i among the last c-r columns of $\boldsymbol{A}$ that is linearly dependent on exactly one of the first r columns of $\boldsymbol{A}$, say, column j, r-1 blocks among the first r blocks in row i of $\boldsymbol{M}$ are all zero. The values in block j in that row are non-negative and total at most one. The values in block j, row j, $j \leq r$, of $\boldsymbol{M}$ sum to one, except when one or more of last c-r rows of $\boldsymbol{M}$ has positive values in block j. In the latter case the values in block j of row j are non-negative, and the total sum of the values in block j, over all the rows of $\boldsymbol{M}$, is one.

It can be shown that the structure of the last c-r blocks of $\boldsymbol{M}$ completes this pattern. For any column i among the last c-r columns of $\boldsymbol{A}$ that is linearly dependent on two or more of the first r columns of $\boldsymbol{A}$, block i in row i of $\boldsymbol{M}$ has non-negative values that add to one, and the other blocks are all zero. In other words, row i has zero values everywhere except block i. For any column i among the last c-r columns of $\boldsymbol{A}$ that is linearly dependent on exactly one of the first r columns of $\boldsymbol{A}$, say, column j, block i in row i of $\boldsymbol{M}$ has non-negative values, so that the total sum of the values in row i, over all the columns of $\boldsymbol{M}$, is one. In this case, in addition, the values in block i of row j are non-negative too, and their total sum is equal to the total sum of the values in block j of row i.

To conclude, $\boldsymbol{M}$ adheres to the structure shown in Figure A.2.

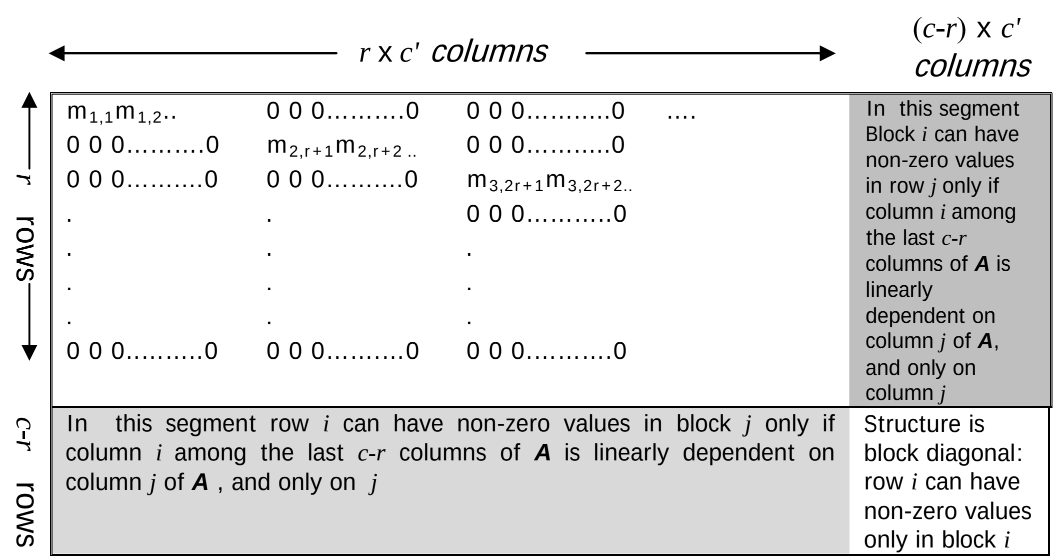


FIGURE A.2 The structure imposed on $\boldsymbol{M}$ (a variation on FIGURE A.1)

Therefore ***G*** has the structure as depicted in Figure 4.1. Therefore, by applying a minor variation on the respective logic in Case 1 above, it can be shown that ***A*** is not GMI than ***G***.

◙

REFERENCES

Abbott, L.F., and Dayan, P. "The effect of Correlated Variability on the Accuracy of a Population Code.", *Neural Computation*, Vol. 11, No. 1, 1999, pp. 91-101.

Ahituv, N. and Ronen, B., "Orthogonal Information Structure - A Model to Evaluate the Information Provided by a Second Opinion." *Decision Sciences*, Vol. 19, No. 2, Spring 1988, pp. 255-268.

Aigner, D.J., "Regression with a Binary Independent Variable Subject to Errors of Observation." *Journal of Econometrics*, Vol. 1, 1973, pp. 49-60.

Ali, K. M., and Pazzani M. J., "Error Reduction through Learning Multiple Descriptions." *Machine Learning*, Vol. 24, No. 3, 1996, pp. 173-202.

Anand, K.S., and Mendelson, H., "Information and Organization for Horizontal Multimarket Coordination." *Management Science*, Vol. 43, No.12, 1997, pp.1609-1627.

Anokhin, P., and Motro, A., "Resolving Inconsistencies in the Multiplex Multidatabase System." Technical Report ISE-TR-99-07. Department of Information and Software Engineering, George Mason University. Fairfax, Virginia. May 2000.

Association of Support Professionals. "Trends in Fee-Based Support." *Association of Support Professionals* (*ASP*), 2002. (Available from www.asponline.com).

Bakos, J.Y.,"Information Links and Electronic Marketplaces: The Role of Interorganizational Information Systems in Vertical Markets." *Journal of Management Information Systems*, Vol. 8, No. 2, 1991, pp. 31-52.

Ballou, D. P. and Pazer, H. L., "Modeling Data and Process Quality in Multi-input, Multi-output Information Systems." *Management Science,* Vol. 31, No. 2, 1985, pp. 150-162

Ballou, D. P., and Pazer, H. L., "Cost/Quality Tradeoffs for Control Procedures in Information Systems." *International Journal of Management Science,* Vol. 15, No. 6, 1987, pp. 509-521.

Ballou, D. P., Pazer, H. L., Belardo, S., and Klein, B., "Implication of Data Quality for Spreadsheet Analysis." *DATA BASE*, Vol. 18, No. 3, 1987, pp. 13-19.

Ballou, D. P., and Tayi, G. K., "Enhancing Data Quality in Data Warehouse Environments." *Communications of the ACM*, Vol. 42, No. 1, 1999, pp. 73-78.

Ballou, D. P., Wang, R. Y., Pazer, H. L., and Tayi, G. K., "Modeling Information Manufacturing Systems to Determine Information Product Quality." *Management Science,* Vol. 44, No. 4, 1998, pp. 462-484.

Bansal, A., Kauffman, R. J. and Weitz, R. R., "Comparing the Modeling Performance of Regression and Neural Networks as Data Quality Varies: A Business Value Approach.", *Journal of Management Information Systems*, Vol. 10, No. 1, 1993, pp. 11 - 32.

Barabash, T. L., "On properties of symbol recognition." *Engineering Cybernetics*, Sept./Oct. 1965, pp. 71-77.

Barua, A., Kriebel, C.H., and Mukhopadhyay, T. "MIS and Information Economics: Augmenting Rich Descriptions with Analytical Rigor in Information Systems Design." *Proceedings of the 10th International Conference on Information Systems* (ICIS), 1989.

Batchelder, W.H., and Romney A.K., "The Statistical Analysis of a General Model for Dichotomous Choice Situations." in Grofman, B., and Owen, G., *Information pooling and group decision making: Proceedings of the Second University of California, Irvine, Conference on Political Economy*, Greenwich, Conn.: JAI Press, 1986.

Black, D., Sanders, S., and Taylor, L., "Measurement of Higher Education in the Census and CPS." 2003 (unpublished paper).

Blackwell, D. "Comparison of Experiments." *Proceedings of the second Berkeley Symposium on Mathematical Statistics and Probability,* University of California Press, 1951, pp. 93-102.

Blackwell, D. "Equivalent comparisons of experiments." *Annals of Mathematical Statistics* Vol. 24, No. 2, 1953, pp. 265-272.

Boehm, B. W., and Sullivan, K. J., "Software economics: A Roadmap." *Proceedings of the 22nd International Conference on Software Engineering* (ICSE), Limerick Ireland, 2000, pp. 319-343.

Bohnenblust, H.F., Shapley, L.S., and Sherman, S., "Reconnaissance in Game Theory." Research Memorandum RM-208, RAND Corporation, Santa Monica, California, 1949.

Boll, C.H., "Comparison of Experiments in the Infinite Case and the Use of Invariance in Establishing Sufficiency." Unpublished P.HD. Thesis, Stanford, 1955.

Bound, J., Brown, C., Duncan, G., and Rodger, W., "Evidence on the Validity of Cross-Sectional and Longitudinal Labor Market Data." *Journal of Labor Economics*, 1994, Vol. 12, No. 3, 1994, pp. 345-68.

Brown, S., Van Bennekom, F., Goffin, K., Hahn, A., and Alexander, J.A. *S-Business: Defining the Services Industry*. Association for Services Management International (AFSMI), October 2001. (Available from www.afsmi.org).

Clemen, R.T., and Winkler, R.L., "Limits for the Precision and Value of Information from Dependent Sources." *Operations Research*, Vol. 33, No. 2, March-April 1985, pp.427-442

Clemons, E.K., and Row, M.C., "Information Technology and Industrial Cooperation: The Changing Economics of Coordination and Ownership." *Journal of Management Information Systems*, Vol.9, No.2, 1992, pp. 9-28.

Cochran, W.G., "On the Performance of the Linear Discriminant Function." *Bulletin of the International Statistics Institute*, Vol. 39, 1962, pp. 435-447.

Condorcet , Nicolas Caritat de *Essai sur l'application de l'analyse a la probabilité des décision rendues à la pluralité des voix.* Paris, 1785.

Cover, T., "The Best Two Independent Measurements are Not the Two Best." *IEEE Transactions on Systems, Man and Cybernetics*, Vol. SMC-4, No. 1, 1974, pp. 116-117.

Cover, T., and Van Campenhout, J., "On the Possible Orderings in the Measurement Selection Problem." *IEEE Transactions on Systems, Man and Cybernetics*, Vol. SMC-7, No. 9, 1977, pp. 657-661.

Cusumano, M.A., "Finding your balance in the products and services debate." *Communications of the ACM*, Vol. 46, No. 3, 2003, pp. 15-17.

Cusumano, M.A., and Selby, R.W., *Microsoft Secrets*. New York, NY: The Free Press, 1995.

Dietterich T.G., "Ensemble Methods in Machine Learning." In J. Kittler and F. Roli, (eds.) *First International Workshop on Multiple Classifier Systems*, Lecture Notes in Computer Science, Springer-Verlag, 2000, pp. 1-15.

Dietterich, T.G., "An Experimental Comparison of Three Methods for constructing ensembles of decision trees: Bagging, Boosting, and Randomization." *Machine Learning*, Vol. 40, No. 2, 2000b, pp. 139-157.

Dover, P.,"Meeting the solutions challenge." *Babson Insight*, 2002.

Elashoff, J.D., Elashoff, R.M., and Goldman, G.E., " On the Choice of Variables in Classification Problems with Dichotomous Variables." *Biometrika*, Vol. 54, 1967, pp. 668-670.

Fang, G.S., "A Note on Optimal Selection of Independent Observables." *IEEE Transactions on Systems, Man and Cybernetics*, Vol. SMC-9, No. 5, 1979, pp. 309-311.

Fayad, M., Laitinen, M., Ward, R.P., "Software Engineering in the Small." *Communications of the ACM*, Vol. 43, No. 3, 2000, pp. 115-118.

Frantsuz, A.G., "Influence of Correlations between Attributes on their Informativeness for Pattern Recognition.", *Engineering Cybernetics*, No 4, July-Aug 1967

Frisch, R., *Statistical Confluence Analysis by Means of Complete Regression Systems*, Publication No. 5, Oslo: University Economics Institute, 1934.

Fuller, W., "Regression Estimation in the Presence of Measurement Errors." Biemer, P.P., Groves, R.M., Lyberg, L.E., Mathiowetz, N.A., Sudman, S. (Eds.) *Measurement Errors in Surveys*, Wiley, 1991.

Garcia-Molina, H., Ullman, J., and Widom, J., *Database Systems: The Complete Book*, Prentice-Hall, 2002.

Goodhue D.L., Wybo M.D. and Kirsh L.J., "The Impact of Data Integration on the Costs and Benefits of Information Systems." *Management Information Systems Quarterly*, Vol.16, No.3; September 1992,pp.293-311.

Greene, W H., *Econometric Analysis* New York: Macmillan, 2000.

Grofman, B., Owen, G., and Feld, S.L., "Thirteen Theorems in Search of the Truth." *Theory and Decision*, Vol. 15, 1983, pp. 261-278.

Gurbaxani, V. and Whang, S., "The Impact of Information Systems on Organizations and Markets." *Communications of the ACM*, Vol. 34, No. 1, 1991, pp. 59-73.

Hansen, L.K., and Salamon, P., "Neural Network Ensembles." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 12, No. 10, 1990, pp. 993-1001.

Hilton, RW "The Determinants of Information Value: Synthesizing Some General Results." *Management Science*, Vol. 27, No. 1, 1981, pp. 57-64

Hussain, A. B. S., "On the Correctness of Some Sequential Classification Schemes in Pattern Recognition." *IEEE Transactions on Computers*, Vol. C-21, 1972, pp. 318-320.

Hwarng, H.B., "Insights into Neural-Network Forecasting of Time Series Corresponding to ARMA(p, q) Structures." *Omega*, Vol. 29, No. 3, 2001, pp. 273-289.

Jacobs, R.A., "Bias/variance Analyses of Mixture-of-Experts Architectures." *Neural computation*, Vol. 9, 1997, pp. 369-383.

Jain, A.K., Duin, R.P.W., and Mao, J., "Statistical Pattern Recognition: A Review." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 1, 2000, pp. 4-37.

Kuncheva, L.I., Whitaker, C.J., Shipp, C.A., and Duin, R.P.W., "Limits on the Majority Vote Accuracy in Classifier Fusion." *Pattern Analysis and Applications,* 6, 2003, pp. 22-31.

Kuncheva, L.I., Whitaker, C.J., "Measures of Diversity in Classifier Ensembles snd their Relationship with the Ensemble Accuracy." *Machine Learning*, Vol. 51, 2003, pp. 181-207.

Kleijnen, J.P.C., *Computers and Profits: Quantifying Financial Benefits of Information*, Reading, Mass.: Addison-Wesley, 1980.

Klein, B.D., and Rossin, D.F., "Data Quality in Linear Regression Models: Effect of Errors in Test Data and Errors in Training Data on Predictive Accuracy." *Informing Science*, Vol. 2, No. 2, 1999.

Klein, B.D., and Rossin, D.F., "Data Quality in Neural Network Models: Effect of Error Rate and Magnitude of Error on Predictive Accuracy." *Omega*, Vol. 27, No. 5, 1999, pp. 569-582

Ladha, K. "Condorcet's Jury Theorem, Free Speech and Correlated Votes." *American Journal of Political Science*, Vol 36, 1992, pp. 617-634.

Ladha, K. "Information Pooling through Majority-Rule Voting: Condorcet's Jury Theorem with Correlated Votes." *Journal of Economic Behavior and Organization*, Vol. 26, 1995, pp. 353-372.

Lee, H.G., Clark T., and Tam, K.Y. "Research Report. Can EDI Benefit Adopters." *Information Systems Research,* Vol. 10, No. 2, June 1999, pp. 186-195.

Liu, Y., and Yao, X. "A cooperative ensemble learning system." *Proceedings of the IEEE International Joint Conference on Neural Networks*, 1998, pp. 2202-2207.

Malone, T. W., Yates, J. and Benjamin, R. I., "Electronic Markets and Electronic Hierarchies." *Communications of the ACM*, Vol. 30, No.6, June 1987, pp. 484-497.

Marschak, J., "Economics of Information Systems." *Journal of American Statistical Association*, Vol 66, No. 333 , March 1971, pp. 192-219.

Marschak, J., and Miyasawa, K., " Economic comparability of information systems." *International Economic Review* Vol. 9, No. 2, 1968, pp. 137-174.

Marschak, J., and Radner, R. *Economic Theory of Teams*, New Haven, Connecticut: Yale University Press, 1972.

McGuire, C. B., "Comperisons of Information Structures." In *Decision and Organization*, C.B. McGuire and R. Radner (Eds.), University of Minnesota Press, 2nd edition, 1986.

McGuire, C.B. and Radner, R. (editors), *Decision and Organization*, University of Minnesota Press, 2nd edition, Minneapolis , Minnesota, 1986 .

Merkhofer, M.W. "The Value of Information Given Decision Flexibility." *Management Science*, 1977.

Mikelson, S. "Should professional services be in your company's services portfolio?" *The Professional Journal 25*, 10, 2001.

Morris, P.A. "Combining Expert Judgements: A Bayesian Approach." *Management Science* , Vol. 23, No. 7, March 1977, pp. 679-693.

Motro, A., and Rakov, I., "Estimating the Quality of Databases." *Proceedings of the International Conference on Flexible Query Answering Systems* (FQAS). Roskilde, Denmark: Springer Verlag, 1998, pp. 298-307.

Mukhopadhyay, T., and Cooper, R.B., "The Impact of Management Information System on Decisions", *OMEGA (International Journal of Management Science)*, Vol. 20, No. 1, 1992, pp. 37-49.

Mukhopadhyay, T., Kekre, S., and Kalathur, S. "Business Value of Information Technology: A study of Electronic Data Interchange." *Management Information Systems Quarterly*, Vol. 19, No. 2, 1995, pp. 137-155.

Nitzan, S. and Paroush, J., "Optimal Decision Rules in Uncertain Dichotomous Choice Situations." *International Economic Review*, Vol. 23, No. 2, June 1982, pp. 289-297.

Nitzan, S. and Paroush, J., "The Significance of Independent Decisions in Uncertain Dichotomous Choice Situations." *Theory and Decision*, Vol. 17, 1984, pp.47-60.

Opdahl, A.L., "A Comparison of Four Families of Multi-Perspective Problem Analysis Methods." *Proceedings of Information Systems Research in Scandinavia* (IRIS 20), Hankø/Norway, Kristin Braa and Eric Monteiro (eds.), 1997.

Radner, R. "Teams." in *Decision and Organization*, C.B. McGuire and R. Radner (Eds.), University of Minnesota Press, 2nd edition, 1986.

Raghunathan, S., "Information Sharing in a Supply Chain: A Note on its Value when Demand is Nonstationary." *Management Science*, Vol. 47, No. 4, 2001, pp. 605-610.

Rakov, I., "Quality of Information in Relational Databases and its Use for Reconciling Inconsistent Answers in Multidatabases." 1997.

Repo, A.J., "The value of information: approaches in economics, accounting, and management science." *Journal of the American Society for Information Science* Vol. 40, No. 2, 1989, pp. 68-85.

Robey, D., "Research Commentary: Diversity in Information Systems Research: Threat, Promise, and Responsibility." *Information Systems Research*, Vol. 7, No. 4, 1996, pp. 400-408.

Rosemann, M., Green, P., "Developing a meta model for the Bunge-Wand-Weber ontological constructs." *Information Systems*, Vol. 27, No. 2, 2002, pp. 75-91.

Sahin, F., and Robinson, E.P., "Flow Coordination and Information Sharing in Supply Chains: Review, Implications, and Directions for Future Research." *Decision Sciences Journal*, Vol. 33, No. 4, 2002, pp. 505-536.

Saleh, S. "Profit metrics for professional services." In *Trends in Fee-Based Support, Association of Support Professionals* (*ASP*), 2002. (Available from www.asponline.com).

Samson, D., Wirth, A., and Rickard, J. "The value of Information from Multiple Sources of Uncertainty in Decision Analysis." *European Journal of Operational Research*, Vol. 39, 1989, pp. 254-260.

Savage, L.J., *The Foundations of Statistics*, Wiley, New York, 1954.

Samuelson, P. A., “The Pure Theory of Public Expenditure.” *The Review of Economics and Statistics*, Vol. 36, No. 4, 1954, pp. 387-389.

Schapire, R. E. “The Boosting Approach to Machine Learning: An overview.” In *MSRI Workshop on Nonlinear Estimation and Classification*, 2002.

Shapiro, C., and Varian, H., *Information Rules*, Boston, MA: Harvard Business School Press, 1998.

Sherman, S., “On the theorem of Hardy, Littlewood, Polya, and Blackwell.” *Proceedings of the National Academy of Sciences*, USA, Vol. 27, 1951, pp. 826-831.

Siau, K., and Rossi, M., “Information Modeling Methods and Methodologies - Evaluation Techniques.” *Proceedings of the 31th Hawaii International Conference on Systems Science* (HICSS), 1998, pp. 312-313.

Simon, C.P., and Blume, L., *Mathematics for Economists*, W.W. Norton & Company, 1994.

Slater, D. Should you join the freeware revolution? *CIO*, March 19, 1999.
Soft Letter. The 2000 Soft Letter 100, *Soft Letter: Trends & Strategies in Software Publishing*. Vol. 16, No. 12, 2000. (Available from www.asponline.com).

Soft Letter & the Association of Support Professionals. Technical Support Cost Ratios. *The Association of Support Professionals*, 1997. (Available from www.asponline.com).

Soft Letter & the Association of Support Professionals. “Technical Support Cost Ratios.” *The Association of Support Professionals*, 2000. (Available from www.asponline.com).

Stallings, W., *Data and Computer Communications*, Prentice Hall, 7th edition, 2003.

Stamper, R., “Management Epistemology: Garbage in, Garbage out.” In Liel B. Methlie and Ralph H. Sprague (Eds.), *Knowledge Representation for Decision Support Systems*, Amsterdam: North-Holland, 1985, pp. 55-80.

Stein, C., “Notes on the Comparison of Experiments.” University of Chicago, 1951.

Toussaint, G.T., “Note on optimal selection of independent binary-valued features for pattern recognition.” *IEEE Transactions on Information Theory*, Vol. IT-17, 1971, p. 618.

Varian, H.R., *Economics of information technology*, Mattioli Lecture, Bocconi University, Milan, Italy, November 2001. (Available from www.sims.berkeley.edu/~hal) von Neumann, J., and Morgenstern, O. *Theory of Games and Economic Behavior*, Princeton Univ. Press, Princeton NJ, 1944.

Wang, R., Storey, V., and Firth, C., "A Framework for Analysis of Data Quality Research." *IEEE Transactions on Knowledge and Data Engineering*, Vol. 7, 1995, pp. 623-640.

Wang, R.Y., and Strong, D.M., "Beyond Accuracy: What Data Quality Means to Data Consumers." *Journal of Management Information Systems*, Vol. 12, 1996, pp. 5-34.

Wand , Y. and Wang , R. Y. , "Anchoring Data Quality Dimensions in Ontological Foundations." , *Communications of the ACM* , Vol. 39 , No. 11 , November 1996 , pp. 86 -95.

Wand, Y., and Weber, R., "An ontological model of an information system." *IEEE Transactions on Software Engineering* (TSE), 1990, pp. 1282–1292.

Wand, Y., and Weber, R., "On the deep structure of information systems." *Information Systems Journal*, Vol. 5, 1995, pp. 203–223.

Wand, Y., Monarchi, D.E., Parsons, J., and Woo, C.C. "Theoretical foundations for conceptual modelling in information systems development." *Decision Support Systems* Vol. 15, 1995, pp. 285-304.

Wilson, B. "Pricing software support." *Annual Hahn Consulting, Inc. Conference* (6th), 2001. (Available from www.hahnconsulting.com).

Winkler, R.L., "Combining Probability Distributions from Dependent Information Sources." *Management Science*, Vol. 27, No. 4, April 1981, pp. 479-488.

Yoon, H., and Sompolinsky, H., "The Effect of Correlations on the Fisher Information of Population Codes." *Advances in Neural Information Processing Systems*, 11. Kearns M.J, Solla S.A and Cohn D.A, Eds., 1999.

Young, P., "Condorcet's Theory of Voting." *American Political Science Review*, Vol. 82, No. 4, 1988, pp. 1231-1244.