ON THE RELATIONSHIPS BETWEEN THE VALUE AND THE AMOUNT OF INFORMATION

By HIROYUKI ITAMI*

1. *Introduction*

The purpose of this paper is to analyze the relationships between the value of information provided by an information system (as developed, for example, in Marschak [10]) and the amount of information provided by an information system (as defined in Lindley [8]). The value is essentially defined as the expected increase of utility by having an information system and the amount is defined as the expected decrease of the degree of uncertainty of the states of nature, the degree of uncertainty of a probability distribution being measured by entropy. The exact expressions for the value and the amount are given below.

Investigations in this paper are particularly concerned with the "consistency" of the rankings of information systems by the value and by the amount. Does an increase in the amount imply an increase in the value? This question seems to be interesting at least on two grounds. First, it asks if an information system is like many commodities in economic analysis for which utility or value is often assumed or considered to be an increasing (or non-decreasing) function of their quantities.[1] Secondly, the "consistency" of rankings will make it possible to use the amount of information (I) as a surrogate of the value of information (V) in information system selection.[2] This possibility has a considerable appeal, at least in a practical sense, because of the practical difficulties of finding a suitable utility function and computing the optimal action and its associated utility which the computation of V requires. The computation of I is simple and free from these difficulties.[3]

In the next section we indicate that there are several similar properties that both V and I have. Section 3 is mostly mathematical and obtains conditions for the two rankings of information systems, one by V and one by I, to be "consistent". In Section 4, we apply the results of Section 3 in a specific context of the investment model with state contingent claims. It is shown that the two rankings of information systems are identical if and only if the utility function is logarithmic. In Section 5, we show that there is a large class of information systems for which the ranking by I can never be reversed by the ranking by V.

---

* Assistant Professor (*Jokyōju*) of Management Science. The author thanks for the helpful comments to the earlier draft of this paper given at Stanford Summer Research Seminar in Accounting.

[1] Of course, under the given definitions of the value and amount of information.

[2] The reader should be reminded that cost aspects of information systems are not considered explicitly in this paper.

[3] For more discussion on the introduction of the concept of the amount of information into the framework of information economics and accounting information evaluation in particular, see Itami [7]. The difficulties mentioned here are especially acute for accounting information evaluation for which, in many instances, neither the utility function to be used and the decision situation to be considered are not given explicitly.

We show this by applying the results of Section 3.

Let us now introduce several notations for the following discussions.

S:	the set of states of nature
s_i:	the i-th state of nature, $i=1, \ldots, n$, an element of S
A:	the set of actions available to the decision maker
a:	an action, an element of A
$U(s, a)$:	utility function of the decision maker
$\varphi(s_i)$:	a prior probability distribution over S held by the decision maker
η:	an information system
Y:	the set of messages from η
y_j:	the j-th message, $j=1, \ldots, m$, an element of Y
$p(y_j/s_i)$:	message generating probability given state s_i. This depends on η and sometimes denoted as $p(y_j/s_i, \eta)$
P:	an information matrix $(n\times m)$, $\{p(y_j/s_i)\}$ of η
$p(y_j)$:	the probability of a message $y_j \cdot \sum_i \varphi(s_i)p(y_j/s_i)$
$\varphi(s_i/y_j)$:	a posterior probability distribution over s_i given a message y_j from η
$a(j)$:	the optimal (expected utility maximizing) action given a message y_j
U^0:	the maximized expected utility under a prior distribution $\varphi(s)$. $\max_a \sum_i \varphi(s_i)U(s_i, a)$
$U(y_j)$:	the maximized expected utility under a posterior distribution $\varphi(s_i/y_j)$. $\max_a \sum_i \varphi(s_i/y_j)U(s_i, a)$
$U(\eta)$:	the expected utility under η. $\sum_j p(y_j)U(y_j)$
$V(\eta)$:	the value of information of η, $V(\eta)=U(\eta)-U^0$
$I(\eta)$:	the amount of information of η, defined below.

To denote the value and amount of information, we use notations like V, $V(\eta)$, $V(P)$, I, $I(\eta)$, $I(P)$ quite freely. $p(\cdot)$ for a probability distribution is also frequently used for different probability distributions as long as there would be no confusion. The following alternative expressions for V and I are also used sometimes in the following analysis.

$$\begin{aligned} V &= \sum_j p(y_j) \max_a \sum_i \varphi(s_i/y_j)U(s_i, a) - U^0 \\ &= \sum_j \max_a \sum_i \varphi(s_i)p(y_j/s_i)U(s_i, a) - U^0 \\ &= \sum_j p(y_j) \sum_i \varphi(s_i/y_j)U(s_i, a(j)) - U^0 \\ &= \sum_i \sum_j \varphi(s_i)p(y_j/s_i)U(s_i, a(j)) - U^0 \end{aligned}$$

$$\begin{aligned} I &= -\sum_i \varphi(s_i) \log \varphi(s_i) + \sum_j p(y_j) \sum_i \varphi(s_i/y_j) \log \varphi(s_i/y_j) \\ &= \sum_i \sum_j \varphi(s_i)p(y_j/s_i) \log \frac{\varphi(s_i/y_j)}{\varphi(s_i)} \end{aligned}$$

2. *Similar Properties of the Value and the Amount of Information*

Both the value and the amount of information depend on the prior distribution φ and the information matrix P. Furthermore, both are defined in terms of the expected difference of certain quantities (i.e. utility for the value and the degree of uncertainty for the amount) between the case with an information system P and the case without it. Therefore, it is perhaps natural to expect some similarities in their properties. Although much cannot be said without specifying the utility function and the decision situation of the decision maker, we shall see that some very basic properties of the value of information are also present in the amount of information. This is somewhat encouraging if one would want to use the amount of information (I) as a surrogate of the value of information (V) in information systems selection.

Now, let us start with an obvious similarity between V and I.

Property 1. Both V and I are non-negative. I is zero if and only if $p(y/s)$ does not depend on s. V is zero if $p(y/s)$ does not depend on s.

For proof, see Lindley [8] and Marschak [10].

Property 2. Both V an I are additive, or

$$V(\eta_1 \cup \eta_2) = V(\eta_1) + V(\eta_2/\eta_1)$$
$$I(\eta_1 \cup \eta_2) = I(\eta_1) + I(\eta_2/\eta_1)$$

Remarks:

Here η_1 and η_2 denote two information systems with the same state space and $\eta_1 \cup \eta_2$ denote the coupled information system in which one receives a pair of messages (y^1, y^2) simultaneously, y^1 from η_1 and y^2 from η_2. $V(\eta_2/\eta_1)$ denote the average additional value of information of η_2 after having received a message from η_1, or

$$V(\eta_2/\eta_1) = \sum_{y^1} p(y^1) V(\eta_2/y^1)$$

where $V(\eta_2/y^1)$ is the additional value of information of having information system η_2 after receiving a message y^1. $V(\eta_2/y^1)$ is computed as usual with $\varphi(s/y^1, \eta_1)$ as a prior and $p(y^2/s, y^1, \eta^2)$ as the message generating probability of η_2. The definition of $I(\eta_2/\eta_1)$ is just the same. It is the average of $I(\eta_2/y^1)$ which is defined similarly as $V(\eta_2/y^1)$.

Proof: By definition

$$V(\eta_2/\eta_1) = \sum_{y^1} p(y^1) U(\eta_2/y^1) - \sum_{y^1} p(y^1) U(y^1)$$
$$V(\eta_1) = \sum_{y^1} p(y^1) U(y^1) - U^0$$

By adding these two together and rewriting $U(\eta_2/y^1)$,

$$\begin{aligned} V(\eta_1) + V(\eta_2/\eta_1) &= \sum_{y^1} \sum_{y^2} p(y^1) p(y^2/y^1) U(y^2/y^1) - U^0 \\ &= \sum_{y^1} \sum_{y^2} p(y^1, y^2) U(y^1, y^2) - U^0 \\ &= V(\eta_1 \cup \eta_2) \end{aligned}$$

The proof for I is similar and given in Lindley [8].

Property 3. Both V and I will never be increased by garbling of an information matrix.

More formally,

$$V(\eta_1) \geq V(\eta_2)$$
$$I(\eta_1) \geq I(\eta_2)$$

if there exists a Markov matrix G such that

$$P_2 = GP_1$$

For proof, see Marschak and Miyasawa [11]. Actually, the existence of garbling is also a necessary condition for $V(\eta_1) \geq V(\eta_2)$ to be true regardless of the utility function and the prior. This is not the case for I.

Property 4. Both V and I are convex functions of P.

Proof: Convexity of I is proved in Lindley [8]. To prove convexity of V, take two information matrices P_1, P_2 and its convex combination $P_3 = \lambda P_1 + (1-\lambda)P_2$, $0 \leq \lambda \leq 1$. Clearly P_3 satisfies the conditions for an information matrix (i.e. Markov matrix).

$$V(P_3) = \sum_y \max_a \sum_s \varphi(s)(\lambda p^1(y/s) + (1-\lambda)p^2(y/s)U(s, a) - U^0$$
$$\leq \sum_y \max_a \lambda \sum_s \varphi(s)p^1(y/s)U(s, a) + \sum_y \max_a (1-\lambda) \sum_s \varphi(s)p^2(y/s)U(s, a) - U^0$$
$$= \lambda V(P_1) + (1-\lambda)V(P_2) \qquad \text{Q.E.D.}$$

An interesting implication of this property would be about the effect of randomization among several alternative information systems. Suppose there are n alternative information systems (P_i) one of which is actually used to emit a message the decision maker receives. Further suppose that he has a prior probability λ_i that the i-th system will be used. Property *4* implies that, both in terms of the value and the amount of information, disclosure of which information system is being used is never worse than non-disclosure on the average basis. Thus, for example, if we consider different accounting alternatives as alternative information systems, disclosure of accounting methods is never worse, on the ex ante basis, than non-disclosure for a single decision maker.

The basic reason why V and I share some basic properties mentioned above is given in the following theorem.

Theorem 1: All the properties which V has regardless of the utility function and the decision situation also hold for I.

It is sufficient to prove that there is at least one special combination of the utility function and the decision situation for which V and I are identical. Section 4 actually gives such an example in the context of investment in risky assets. This theorem is a strong indication of similarity of the properties of V and I as functions of an information matrix and a prior distribution. Note that the theorem holds not only for the four properties mentioned in this paper, but also to any property that V has independent of the utility function and the decision situation.

3. *Conditions for Consistency of Rankings by* I *and* V

An information system η which can be abstractly represented by an information matrix P may be considered to consist of several basic information system elements or inputs which can be varied by an information system designer. Concretely, he does not change the message generating probability itself but manipulates some design parameters which eventually affect P. For example, in a sampling information system for statistical estimation the sample size and the sample precision are most likely design parameters. Then, in this light, the question of the consistency of ranking of information systems by V and I becomes a question of the consistency of the effects on V and I of changing design parameters.

Now, supposing an information system is made up with K design parameters, $\varepsilon_k(k=1, \ldots, K)$, let us assume that every $p(y_j/s_i)$ is a differentiable function of ε_k's.[4] Then, I is differentiable with respect to ε_k's. For the differentiability of V, we first assume that $U(s, a)$ is a differentiable function of a H-dimentional vector a. The constraint set of a is also assumed to be $A=\{a/t_l(a)\leq 0, l=1, \ldots, L)$ where $t_l(a)$ is a differentiable function for all l. Then, the optimality condition (Kuhn-Tucker condition) of the decision maker's problem after receiving a message y_j is

$$(1)\quad \sum_i \frac{\partial U}{\partial a_h}\varphi(s_i/y_j)=\sum_l \lambda_l \frac{\partial t_l}{\partial a_h} \quad \text{for } h=1, \ldots, H$$

$$(2)\quad \lambda_l t_l(a)=0 \quad \text{for } l=1, \ldots, L$$

$$(3)\quad t_l(a)\leq 0 \quad \text{for } l=1, \ldots, L$$

$$(4)\quad \lambda_l \geq 0 \quad \text{for } l=1, \ldots, L$$

Here λ_l is a Lagrangean multiplier.

Now, we have the following lemma for the partial derivatives of V and I. To avoid being entangled with unnecessary mathematical subtleties, we just assume in the following a(j) is differentiable with respect to ε_k's, where $a(j)$ is the optimal decision after receiving a message y_j. This assumption ensures the differentiability of V.

Lemma 1:

$$(5)\quad \frac{\partial V}{\partial \varepsilon_k}=\sum_i \sum_j \varphi(s_i)\frac{\partial p(y_j/s_i)}{\partial \varepsilon_k}U(s_i, a(j))$$

$$(6)\quad \frac{\partial I}{\partial \varepsilon_k}=\sum_i \sum_j \varphi(s_i)\frac{\partial p(y_j/s_i)}{\partial \varepsilon_k}\log\frac{\varphi(s_i/y_j)}{\varphi(s_i)}$$

Proof: Since $V=\sum_j p(y_j)\sum_i U(s_i, a(j))\varphi(s_i/y_j)-U^0$

$$\frac{\partial V}{\partial \varepsilon_k}=\sum_j\sum_i \frac{\partial}{\partial \varepsilon_k}(p(y_j)\varphi(s_i/y_j))U(s_i, a(j))+\sum_j\sum_i p(y_j)\varphi(s_i/y_j)\sum_h \frac{\partial U}{\partial a_h}\frac{\partial a_h(j)}{\partial \varepsilon_k}$$

$$=\sum_i\sum_j \varphi(s_i)\frac{\partial p(y_j/s_i)}{\partial \varepsilon_k}U(s_i, a(j))+\sum_j p(y_j)\sum_h \frac{\partial a_h(j)}{\partial \varepsilon_k}\sum_l \lambda_l \frac{\partial t_l}{\partial a_h}$$

Since either $\lambda_l=0$ or $t_l(a(j))=0$ from (2),

$$\lambda_l \cdot \sum_h \frac{\partial t_l}{\partial a_h}\cdot\frac{\partial a_h(j)}{\partial \varepsilon_k}=0$$

Therefore, we obtain (5).

Since $I=\sum_i\sum_j \varphi(s_i)p(y_j/s_i)\log\frac{\varphi(s_i/y_j)}{\varphi(s_i)}$,

$$\frac{\partial I}{\partial \varepsilon_k}=\sum_i\sum_j \varphi(s_i)\frac{\partial p(y_j/s_i)}{\partial \varepsilon_k}\log\frac{\varphi(s_i/y_j)}{\varphi(s_i)}+\sum_i\sum_j \varphi(s_i)p(y_j)\frac{\partial \varphi(s_i/y_j)}{\partial \varepsilon_k}$$

Furthermore,

[4] Since ε_k can be related to $p(y_j/s_i)$ in any way, the differentiability condition should not be restrictive.

$$\frac{\partial \varphi(s_i/y_j)}{\partial \mathcal{E}_k}=\frac{\frac{\partial p(y_j/s_i)}{\partial \mathcal{E}_k}p(y_j)-p(y_j/s_i)\sum_i \varphi(s_i)\frac{\partial p(y_j/s_i)}{\partial \mathcal{E}_k}}{p(y_j)^2}$$

Therefore,

$$\sum_i \sum_j \varphi(s_i)p(y_j)\frac{\partial \varphi(s_i/y_j)}{\partial \mathcal{E}_k}=\sum_i \sum_j \left\{\varphi(s_i)\frac{\partial p(y_j/s_i)}{\partial \mathcal{E}_k}-\varphi(s_i/y_j)\sum_i \varphi(s_i)\frac{\partial p(y_j/s_i)}{\partial \mathcal{E}_k}\right\}$$

$$=\sum_i \sum_j \varphi(s_i)\frac{\partial p(y_j/s_i)}{\partial \mathcal{E}_k}-\sum_j \sum_i \varphi(s_i)\frac{\partial p(y_j/s_i)}{\partial \mathcal{E}_k}\sum_i \varphi(s_i/y_j)=0$$

Thus, we obtain (6). Q.E.D.

Using this lemma, we can now derive a general condition for V and I to move in the same direction as $\mathcal{E}_k$'s change. That is, denoting the total differentials of V and I by dV and dI, the necessary and sufficient condition for dI and dV to be of the same sign or both to be zero.

Theorem 2: The following two sets of statements are equivalent.

(i) $dI>0 \Rightarrow dV>0$, $dI=0 \Rightarrow dV=0$ and $dI<0 \Rightarrow dV<0$

(ii) The following K equations system (T) of u_1, u_2 and w_i $(i=1, \ldots, n)$ has a solution in which $u_1>0$ and $u_2>0$.

$$(T)\quad \sum_i \sum_j \varphi(s_i)\frac{\partial p(y_j/s_i)}{\partial \mathcal{E}_k}(u_1 U(s_i, a(j))-u_2 \log \varphi(s_i/y_j)+w_i)=0 \qquad \text{for } k=1, \ldots, K$$

Proof: By definition and Lemma 1

$$(7)\quad dV=\sum_k \frac{\partial V}{\partial \mathcal{E}_k}d\mathcal{E}_k=\sum_k \sum_i \sum_j \varphi(s_i)\frac{\partial p(y_j/s_i)}{\partial \mathcal{E}_k}U(s_i, a(j))d\mathcal{E}_k$$

$$(8)\quad dI=\sum_k \frac{\partial I}{\partial \mathcal{E}_k}d\mathcal{E}_k=\sum_k \sum_i \sum_j \varphi(s_i)\frac{\partial p(y_j/s_i)}{\partial \mathcal{E}_k}\log \frac{\varphi(s_i/y_j)}{\varphi(s_i)}d\mathcal{E}_k$$

We have one condition on $d\mathcal{E}_k$'s to keep P as a Markov matrix.

$$(9)\quad \sum_j \sum_k \frac{\partial p(y_j/s_i)}{\partial \mathcal{E}_k}\partial \mathcal{E}_k=0 \qquad \text{for all } i$$

Now the statement (i) is equivalent to saying that none of the following four systems of inequalities of $d\mathcal{E}_k$'s have a solution under the condition (9).

(T_1) $dI\leq 0$ and $dV>0$

(T_2) $dI>0$ and $dV\leq 0$

(T_3) $dI=0$ and $dV<0$

(T_4) $dI<0$ and $dV=0$

Using Motzkin's theorem of the alternative in Mangasarian [9, p. 28], the infeasibility of (T_1) under (9) is equivalent to the feasibility of

$$\sum_i \sum_j (u_1\varphi(s_i)U(s_i, a(j))-u_2\varphi(s_i)\log \frac{\varphi(s_i/y_j)}{\varphi(s_i)}+z_i)\frac{\partial p(y_j/s_i)}{\partial \mathcal{E}_k}=0 \quad \text{for all } k$$

in u_1, u_2, z_i $(i=1, \ldots, n)$ with $u_1>0$, $u_2\geq 0$. Since z_i can depend on anything but y_j without any restriction, the feasibility of the above equation system is equivalent to the feasibility of (T) with $u_1>0$ and $u_2\geq 0$.

By developing similar equivalence relations for the infeasibility of (T_2), (T_3) and (T_4) through Motzkin's theorem, we obtain the desired result. Q.E.D.

The above theorem gives us the necessary and sufficient condition for a local isomorphism between the changes in I and V due to varying design parameters, ε_k's. Extending this into the conditions for global identity of rankings of information systems (or information system changes under condideration manifest in $\frac{\partial p(y_j/s_i)}{\partial \varepsilon_k}$), we have the following theorem:[5]

Theorem 3: Given U, A, φ and $p(y_j/s_i)$ as functions of ε_k's, for the rankings of information systems by V and I to be identical the following two conditions are necessary and sufficient.

(i) V is a function of I,

(ii) Condition (ii) of Theorem 2 holds.

Proof: Since V has to be a strictly increasing function of I for the two rankings to be identical and vice versa, the theorem easily follows.

The condition (i) requires that I is a finer function of P and φ than V, which is a rather strong requirement. There does not seem to exist any easy way to check (i). An obvious sufficient condition is that I as a function of ε_k's has an inverse function.

Since a requirement of the global identity of two rankings is rather strong, one may well be content with a milder requirement that the ranking by I will never be reversed by the ranking by V, or $I(\eta_1)>I(\eta_2)\Rightarrow V(\eta_1)\geq V(\eta_2)$. Contrary to the requirement of identical rankings, we allow the case $V(\eta_1)=V(\eta_2)$ when $I(\eta_1)>I(\eta_2)$. We also allow η_1 and η_2 can be ranked in any way when $I(\eta_1)=I(\eta_2)$. This irreversibility requirement is especially meaningful if one would want to use I as a surrogate criterion of information system choice when a utility function is not clearly known. The following theorem gives us a necessary condition for global irreversibility of the ranking by I and the necessary and sufficient condition for local irreversibility. By local irreversibility of the ranking by I, we mean $dI>0\Rightarrow dV\geq 0$ and $dI<0\Rightarrow dV\leq 0$ for a small change of ε_k's.

Theorem 4: The following condition is the necessary and sufficient condition for $dI>0\Rightarrow dV\geq 0$ and $dI<0\Rightarrow dV\leq 0$, and thus a necessary condition for the ranking by I to irreversible globally by the ranking by V.

(i) The equations system (T) of u_1, u_2 and w_i $(i=1, \ldots, n)$ has a sloution in which $u_1\geq 0$, $u_2\geq 0$ and $u_1+u_2>0$.

Proof: Since $dI>0\Rightarrow dV\geq 0$ and $dI<0\Rightarrow dV\leq 0$ is equivalent to the infeasibility of

(T_5) $dI>0$ and $dV<0$

(T_6) $dI<0$ and $dV>0$

under the condition (9), we can obtain the desired result by applying Motzkin's theorem of the alternative as we did in Theorem 2. Q.E.D.

A simple sufficient condition to make the condition (i) in Theorem 4 the necessary *and* sufficient condition for global irreversibility is that we restrict ourselves for such a range and movement of ε_k's for which $dI\geq 0$. We formally state this as a corollary because it will be used in Section 5.

Corollary to Theorem 4:

If $dI\geq 0$ for a range and movement of ε_k's we are interested in, then the condition (i) of

[5] Here, by "the identity of rankings", we exclude such cases in which η_1 and η_2 are ranked same by V but differently by I. We are concerned with strict identity of two rankings.

Theorem 4 is necessary and sufficient condition for the ranking by I to be irreversible by the ranking by V for the range and movement of ε_k's.

Theorem 3 and 4 are general results in the sense that they can be used to check the identity or irreversibility of rankings for any given U, A, φ and P. They can also be used to derive conditions for identical or irreversible rankings on

(a) classes of utility functions and decision situations when no or weak assumptions are made for information system changes,
(b) classes of information systems (or their changes) when no or weak assumptions are made on U, A and others.

In the following analysis, both questions (a) and (b) will be discussed. In the rest of this section and the next section, we will try to nail down the classes of the utility functions (and the decision situations) for which we can be assured of the identical rankings by V and I for an arbitrary set of information matrices. The next section treats this problem in a specific context of securities investment (or more generally exchange economy under uncertainty). Concerning (b), Section 5 shows that for a certain class of information systems, the irreversibility of the ranking by I exists under a mild assumption on the utility function and the decision situation.

A requirement of identical rankings by V and I for any information system changes or any set of alternative information systems is very strong and is likely to be satisfied only for a rather restricted class of utility functions and decision situations. The following theorem proves that this intuition is correct, but also shows some interesting cases are included in this restricted class of utility functions and decision situations.

Theorem 5: The necessary and sufficient condition for the statement (ii) of Theorem 2, and thus a necessary condition for the identical rankings of information systems by V and I for any set of information systems, is that the utility function and the decision situation are such that

(10) $U(s_i, a(j))=\alpha \log \varphi(s_i/y_j)+f(s_i)$ for all i and j

where α is a positive number which does not depend on either s_i nor y_j and $f(s_i)$ is an arbitrary function of s_i which does not depend on y_j.

Remarks: Note that α and $f(s_i)$ may depend on the value of design parameters as they change.

Proof: From the system (T), it is easy to see that for (T) to hold for any $\dfrac{\partial p(y_j/s_i)}{\partial \varepsilon_k}$, it is necessary and sufficient that the terms in the parenthesis on the left-hand are identically zero for all i, j and k. Thus (10) is obtained. Due to Theorem 3, this is a necessary condition for the identical rankings by V and I. Q.E.D.

A special case of interest which satisfies (10) is a securities investment model of Arrow-Debreu type contingent claims with logarithmic utility of income. It will be treated in depth in the next section. Another case of interest in which (10) holds is the case of quadratic loss estimation of the mean of normal distribution with normal prior.[6] Moreover, in this special case, V can be shown to be a function of I and therefore (10) becomes the necessary and

[6] For example, see DeGroot [4, Ch. 11]. Also, see Itami [7] for a detailed analysis of the relationships between V and I in this case. Although the results in this paper are developed in the discrete probability case, extension to the continuous case is immediate.

sufficient condition for identical rankings.

4. *Identical Rankings in the Investment Model with State Contingent Claims*

In this section, we analyze the question of the identity of rankings of information systems by V and by I within the framework of the investment model with state contingent claims. The model was first proposed by Arrow [1] and is now considered one of very general and basic frameworks of a pure exchange economy under uncertainty and investment in risky assets.[7] Therefore, analysis of the identical ranking question in this particular framework is especially meaningful for accounting information evaluation. The relationship between V and I in this framework was first analyzed by Arrow [2].

Notationally, let x_i be the odds on the occurrence of state s_i or the rate of return of the i-th state contingency claim. Let a_i be the proportion of an investor's fixed total wealth to be bet on the i-th state contingency claim. The investor's decision model, when the probability distribution of states is, say $\varphi(s_i)$, becomes

$$\max \sum_i \varphi(s_i) U(a_i x_i)$$
$$\text{s.t. } \sum_i a_i \leq 1,\ a_i \geq 0$$

Obviously, the role of an information system in this framework is to help the investor revise the probability of occurrence of state s_i, however it may be defined in any particular case.[8]

The major result of this section is presented in Theorem 6 below. It says, in short, that the rankings of information systems by V and I are identical over an arbitrary set of information systems *if and only if* the utility of income is of logarithmic form.

Theorem 6: Assume that an investor's utility function $U(w)$ of income (w) is strictly increasing, twice differentiable, and its second derivative never vanishes. Assume further that an investor invests all the money in contingent claims and there will always be positive investment for all the state contingent claims at optimum. Then for the rankings of information systems by V and I to be identical for any set of alternative information systems, it is necessary and sufficient that $U(w)$ is of the form

(11) $U(w) = \log(w+\beta),\ \beta \geq 0.$

Remarks: For an investor to invest all the money (i.e., $\sum_i a_i = 1$), it is sufficient that there exists a sure system of bets (i.e., $\sum_i \frac{1}{x_i} \leq 1$). See Arrow [2]. Also note that β in (11) is not restricted in any way except $\beta \geq 0$.

Proof: The Kuhn-Tucker condition for optimality for the investor's problem after receiving a message y_j are, under the assumptions of no savings and positive investment for each claim,

(12) $\varphi(s_i/y_j) x_i U'(a_i(j) x_i) = \lambda(j)$ for all i

(13) $\sum_i a_i(j) = 1,$

[7] The model is also known as the states of the world model. See, for example, Fama and Miller [5].

[8] For a view which considers one of the major roles of financial statements as such for a richly definable *states*, see Beaver [3].

where $a_i(j)$ is the optimal solution and $\lambda(j)$ is a Lagrange multiplier. Let us first prove the sufficiency of (11). From (12), it is easy to see

(14) $\varphi(s_i/y_j)x_i=\lambda(j)(a_i(j)x_i+\beta)$.

Therefore,

$$\log(a_i(j)x_i+\beta)=\log\varphi(s_i/y_j)+\log x_i-\log\lambda(j)$$

and,

$$\sum_i \varphi(s_i/y\,)U(a_i(j)x_i)=-\log\lambda(j)+\sum_i \varphi(s_i/y_j)\log x_i+\sum_i \varphi(s_i/y_j)\log\varphi(s_i/y_j)$$

Similarly, for the case without an information system,

$$U^0=-\log\lambda^0+\sum_i \varphi(s_i)\log x_i+\sum_i \varphi(s_i)\log\varphi(s_i).$$

Thus, the value of information is

$$V=-\sum_j p(y_j)\log\lambda(j)+\log\lambda^0+I.$$

Actually, $\lambda(j)$ does not depend on j under our assumptions. We can see that by dividing (14) by x_i and summing over i,

$$1=\lambda(j)+\lambda(j)\beta\sum_i \frac{1}{x_i},$$

which gives us $\lambda(i)$ independent of j. Thus, V is identical with I.

To prove the necessity part, first observe from Theorem 5 that it is necessary

(15) $U(a_i(j)x_i)=\alpha\log\varphi(s_i/y_j)+f(s_i),\ \alpha>0$

Substituting (12) into (15), we obtain

(16) $U(a_i(j)x_i)=-\alpha\log U'(a_i(j)x_i)+\alpha\log\lambda(j)-\alpha\log x_i+f(s_i)$

Let $z=p(y_j)$ for some j and differentiate (16) partially with respect to z. Since α and $f(s_i)$ do not depend on j,

(17) $\left\{U'(a_i(j)x_i)+\alpha\dfrac{U''(a_i(j)x_i)}{U'(a_i(j)x_i)}\right\}x_i\dfrac{\partial a_i(j)}{\partial z}=\alpha\dfrac{1}{\lambda(j)}\dfrac{\partial\lambda(j)}{\partial z}$

Now, for any i there exists such $a_i(j)$ as a differentiable function of z so that $\lambda(j)$ remains constant in (12) as z changes locally. This is seen by applying the implicit function theorem to (12) keeping $\lambda(j)$ constant. Namely, (12) with constant $\lambda(j)$ defines an implicit function of z and $a_i(j)$. From this implicit function, it is possible to find a differentiable function $a_i(j)$ of z if the partial derivative of LHS of (12) with respect to $a_i(j)$ is continuous in the neighborhood and does not vanish. This partial derivative is

$$\varphi(s_i/y_j)x_iU''(a_i(j)x_i)x_i$$

and this is continuous and does not vanish because of our assumptions. Thus, for some changes of z, $\dfrac{\partial\lambda(j)}{\partial z}=0$.

Furthermore we have, from (12)

(18) $\dfrac{\partial a_i(j)}{\partial z}=-\dfrac{\dfrac{\partial\varphi(s_i/y_j)}{\partial z}x_iU'(a_i(j)x_i)}{\varphi(s_i/y_j)x_i^2U''(a_i(j)x_i)}$

Since $U(w)$ is assumed to be strictly increasing, (18) does not vanish. Thus, from (17) we see that for such a function $a_i(j)$ of z with $\dfrac{\partial\lambda(j)}{\partial z}=0$ and $\dfrac{\partial a_i(j)}{\partial z}\neq 0$,

$$\alpha U''(a_i(j)x_i)+\{U'(a_i(j)x_i)\}^2=0$$

Since we are concerned with freely changing information systems, this implies that $U(w)$ has to satisfy the differential equation of the following form.

(19) $\alpha U''(w)+\{U'(w)\}^2=0 \quad$ for $w>0$

Solving this with a condition $U'(w)>0$ for $w>0$, we obtain

$$U(w)=\log(\alpha w+\beta),\ \beta\geq 0.$$

It is clear that letting $\alpha=1$ is not a restriciton for $U(w)$ and we obtain (11). Q.E.D.

The sufficiency part of the proof was essentially indicated in Arrow [2]. The necessity part is a new result which seems to clarify the implication of using I as a criterion of information systems selection, especially for financial accounting information systems whose major function is to provide information to the investors in securities markets. It means that to use I as an information choice criterion and rely on it completely implies the logarithmic utility function. The fact that only one specific utility function is compatible with the use of I as a surrogate has certainly a negative connotation. Yet, the specific form, logarithmic, is somewhat encouraging because it is one of the simplest utility functions to display several intuitively appealing properties of risk aversion.[9] Furthermore, Hakansson [6] shows in a multiperiod portfolio choice problem with a capital growth objective that in order to behave optimally in the long run the investor should act at each decision point *as if* he had a short-run logarithmic utility function of the end of period wealth. Therefore, the static model with the logarithmic utility function, which is implied by the use of I for information choice, is quite appropriate in rather general circumstances.

5. *A Case of Irreversible Ranking by* I: *A Class of Information Systems*

In this section, we shall apply Theorem 4 and its Corollary developed in Section 3 and identify a class of information systems for which the ranking by I cannot be reversed by the ranking by V under a mild condition. First, let us start by describing the kind of information system changes by which any two information systems in this class are related to each other.

Consider a change in an information matrix that increases at most one element of each row and decreases the same number (zero or one) of elements of the same row. That is, for each row i of P, we consider an increase of one $p(y_j/s_i)$ and the associated decrease of another $p(y_j/s_i)$ for the same i. Without loss of generality we can consider such changes as a result of an increase of a single design parameter, say ε_k. Denoting the index j of the increased $p(y_j/s_i)$ for each i by $j(i)$ and that of the decreased $p(y_j/s_i)$ by $h(i)$, suppose further that for each i

$$\varphi(s_i/y_{j(i)})\geq\varphi(s_i/y_{h(i)}).$$

Or we consider such $\dfrac{\partial p(y_j/s_i)}{\partial\varepsilon_k}$ that for each i

(20) $\dfrac{\partial p(y_{j(i)}/s_i)}{\partial\varepsilon_k}>0$

[9] See Pratt [12].

(21) $\dfrac{\partial p(y_{h(i)}/s_i)}{\partial \mathcal{E}_k}<0$

for such $j(i)$ and $h(i)$ that

(22) $\varphi(s_i/y_{j(i)})\geq\varphi(s_i/y_{h(i)})$.

It is easy to see that for this kind of change in an information matrix, $\varphi(s_i/y_{j(i)})$ will increase and $\varphi(s_i/y_{h(i)})$ will decrease for each state i. Thus, it makes the divergence between $\varphi(s_i/y_{j(i)})$ and $\varphi(s_i/y_{h(i)})$ all the more greater. It also makes the posterior probability of the state i carry more weight than before in the posterior distribution given the message $j(i)$, collecting this increased weight from the probability of other states. It also makes the posterior probability of the state i carry less weight than before in the posterior distribution given the message $h(i)$, spreading this reduced weight over other states. Thus, in a rough sense, the above kind of pair-wise changes makes the posterior distribution after receiving messages more different from each other and it makes the state i more associated with the message $j(i)$ and less associated with the message $h(i)$. Since the obviously least meaningful information system is the one which gives the same posterior distribution no matter what message is received, we may be somewhat justified to say that after the above kind of pair-wise changes the information system is more meaningful or "more accurate". This intuitive argument of "more accuracy" of an information system due to the above kind of change is supported by the following lemma. It is shown there that the change will never decrease the amount of information of an information system. For this and the above intuitive reasons, we say that one information system η_1 is preferred to another system η_2 by the Accuracy-ordering (or AC-ordering) if η_2 can be constructed from η_1 by a series of pair-wise changes of the avove kind. Alternatively, we may say η_1 is AC-preferred to η_2.

Lemma 2: Consider $\dfrac{\partial p(y_i/s_i)}{\partial \mathcal{E}_k}$ of the above kind. Then, for such information system changes,

(23) $\dfrac{\partial I}{\partial \mathcal{E}_k}\geq 0$

Proof: Since we are considering a change of a single parameter only, the condition (9) reduces to, in our case here,

(24) $\dfrac{\partial p(y_{j(i)}/s_i)}{\partial \mathcal{E}_k}=-\dfrac{\partial p(y_{h(i)}/s_i)}{\partial \mathcal{E}_k}>0$ for each i

Then, from (6) of Lemma 1, we have

(25) $\dfrac{\partial I}{\partial \mathcal{E}_k}=\sum_i \varphi(s_i)\dfrac{\partial p(y_{j(i)}/s_i)}{\partial \mathcal{E}_k}\log\dfrac{\varphi(s_i/y_{j(i)})}{\varphi(s_i/y_{h(i)})}$

Due to the assumption of the changes, we obtain (23). Q.E.D.

Effectively, this lemma says that for a set of information systems which can be ordered by the AC-ordering, the ranking by I will never reverse the ranking by the AC-ordering. Note that like the ordering of information systems by "garbling" (Property 3 of Section 2), the AC-ordering here is a partial ordering of information systems. Unlike the ordering by "garbling", the AC-ordering depends on the prior distribution.

For a set of information systems which can be ordered by "garbling", Property 3 implies that the ranking by I can never be reversed by the ranking V. This is true, under a mild condition, for a set of information systems which can be ordered by the AC-ordering given a prior $\varphi(s)$. Theorem 7 proves this, thus identifying another class of information systems for which the ranking by I can never be reversed by the ranking by V.

Theorem 7: Assume that the utility function and the decision situation is such that for two different probability distribution of s_i, $\varphi'(s_i)$ and $\varphi''(s_i)$, the corresponding expected utility maximizing actions, a' and a'' respectively, have a property

(26) $U(s_i, a') \geq U(s_i, a'')$

for all i such that $\varphi'(s_i) > \varphi''(s_i)$.

Then, given U, A, and $\varphi(s)$, the ranking of information systems by I will never be reversed by the ranking by V for any set of information systems which can be ordered by the AC-ordering.

Remarks: Assumption (26) of this theorem may be considered a condition on the sensitivity of the optimal action. Greater probability of s_i has to be accompanied by such a new optimal sction that gives no less utility when s_i actually occurs. This seems to be a mild condition. Note that (26) is satisfied if $a'=a''$.

Proof: By definition, between any two information systems in the set of interest here, we can consider a series of design parameter changes as in (20). Take any ε_k in that series and we prove for the $\frac{p(y_j/s_i)}{\partial \varepsilon_k}$, the equations system ($T$) has a solution in which $u_1 \geq 0$, $u_2 \geq 0$ and $u_1 + u_2 > 0$. First, observing (24), the equations system (T) becomes

$$(27) \quad u_1 \sum_i \varphi(s_i) \frac{\partial p(y_{j(i)}/s_i)}{\partial \varepsilon_k} - \{U(s_i, a(j(i))) - U(s_i, a(h(i)))\}$$

$$-u_2 \sum \varphi(s_i) \frac{\partial p(y_{j(i)}/s_i)}{\partial \varepsilon_k} \log \frac{\varphi(s_i/y_{j(i)})}{\varphi(s_i/y_{h(i)})} = 0$$

Suppose $\varphi(s_i/y_{j(i)}) > \varphi(s_i/y_{h(i)})$. The coefficient of u_1 is non-negative and the coefficient of u_2 is positive. Then we can have a solution to (27) such that $u_1 > 0$, $u_2 \geq 0$. Now suppose $\varphi(s_i/y_{j(i)}) = \varphi(s_i/y_{h(i)})$. Since the coefficient of u_2 is now zero, we can have a solution $u_1 = 0$, $u_2 > 0$ to (27). Since $dI \geq 0$ for an increase of an ε_k as proved in Lemma 2, Corollary to Theorem 4 gives us the desired result. Q.E.D.

REFERENCES

1. Arrow, Kenneth J., "The Role of Securities in the Optimal Allocation of Risk-bearing," *Review of Economic Studies*, April 1964, pp. 91–96.
2. ———, "The Value of and Demand for Information", Ch. 5 in *Decision and Organization*, ed. C.B. McGuire and R. Radner, Amsterdam: North-Holland Publishing Co., 1972, pp. 131–140.
3. Beaver, William H., "Financial Statements: Issues of Preparation and Interpretation", unpublished, Graduate School of Business, Stanford University, 1975.
4. DeGroot, Morris H., *Optimal Statistical Decisions*, New York: McGraw-Hill, 1970.
5. Fama, Eugene F., and Merton H. Miller, *The Theory of Finance*, New York: Holt, Rinehart and Winston, 1972.
6. Hakansson, Nils H., "Capital Growth and the Mean-Variance Approach to Portfolio Selection", *Journal of Financial and Quantitative Analysis*, January 1971, pp. 517–558.

7. Itami, Hiroyuki, "The Concept of the Amount of Information in Information Evaluation", *Hitotsubashi Journal of Commerce and Management*, 1977, pp. 33–43.
8. Lindley, D.V., "On a Measure of the Information Provided by an Experiment", *Annals of Mathematical Statistics*, Vol. 27, 1956, pp. 986–1005.
9. Mangasarian, Olvi L., *Nonlinear Programming*, New York: McGraw-Hill, 1969.
10. Marschak, Jacob, "Economics of Information Systems", *Journal of the American Statistical Association*, March 1971, pp. 192–219.
11. ——— and Koichi Miyasawa, "Economic Comparability of Information Systems", *International Economic Review*, June 1968, pp. 137–174.
12. Pratt, John W., "Risk Aversion in the Small and in the Large", *Econometrica*, January–April 1964, pp. 122–136.