

THE VALUE OF A STOCHASTIC INFORMATION STRUCTURE

YARON AZRIELI AND EHUD LEHRER^{†*}

June 5, 2005

ABSTRACT. Upon observing a signal, a Bayesian decision maker updates her probability distribution over the state space, chooses an action, and receives a payoff that depends on the state and the action taken. An information structure determines the set of possible signals and the probability of each signal given a state. For a fixed decision problem (consisting of a state space, action set and utility function) the value of an information structure is the maximal expected utility that the decision maker can get when the observed signals are governed by this structure.

A decision problem induces a preference order over information structures: an information structure I is preferred over I' whenever its value (that corresponds to the decision problem) is higher than that of I' . We provide necessary and sufficient conditions for preference orders over information structures to be induced by a decision problem. We also characterize the functions defined over information structures that measure their value. It turns out that three conditions play a major role in the characterization of these functions: additive separability, convexity and continuity.

[†]School of Mathematical Sciences, Tel Aviv University, Tel Aviv 69978, Israel.
e-mail: azrieliy@post.tau.ac.il, lehrer@post.tau.ac.il.

*An anonymous referee of *Games and Economic Behavior* is particularly acknowledged for his/her insightful comments. We also thank Dinah Rosenberg, Eran Shmaya and the associate editor of *Games and Economic Behavior* for their helpful comments.

1. INTRODUCTION

Comparison of different information structures¹ has been a matter of interest for many years. In a seminal work, Blackwell (1951, 1953) suggested that information structures will be ordered according to the expected utility they yield for the decision maker (DM). Such ordering depends on the particular decision problem that the DM is facing. Indeed, it is possible that one information structure is better than another in a certain problem, but when the problem changes the previously better structure becomes the worse.

The approach taken by Blackwell (1951, 1953) is the following. One information structure, say I , is said to be ‘*better than*’ another, (I') , if *whatever the decision problem is*, the expected utility of the DM is higher when the information structure is I . This definition induces a partial order over information structures. Blackwell showed that this order can be equivalently defined in several different ways, some of them purely probabilistic.

Blackwell’s work initiated a vast literature in this direction². While we consider stochastic information structures similar to that treated by Blackwell, the questions we pose are of a different nature.

A decision problem is defined by a state space, a prior distribution, an action set and a utility function. Before taking a decision, the DM receives a stochastic signal that depends on the state realized. The set of possible signals and the probability of each signal given the state are determined by the particular information structure of the case. Upon receiving a signal the DM updates her belief and takes an action that maximizes her expected utility. For a *given* decision problem, different information structures determine potentially different maximal-achievable expected utilities.

When the DM is asked to choose from two information structures, she prefers the one that yields the higher expected payoff. Thus, a decision problem induces a complete order over information structures. The question arises as to what preference orders over information structures are consistent with an expected utility maximization. That is, what preference orders over information structures are induced by *some* decision problem.

In order to answer this question we analyze *information functions*. The value of an information structure is the maximal achievable expected utility that corresponds to it. The information function is the real-valued function that attaches to any information structure its value. The information function can be interpreted as a summary of the data collected by an outside

¹In the statistics literature, information structures are commonly referred to as *statistical experiments*.

²For a comprehensive survey of this literature see Torgersen (1991). A shorter review can be found in Le Cam (1996).

observer about the payoffs the DM received after having obtained information on the prevailing state through various structures.

Gilboa and Lehrer (1991) investigated properties of information functions whose domain is restricted to the set of deterministic information structures. In such information structures the signal observed by the DM is uniquely determined by the state of nature. Therefore, every deterministic information structure generates a partition of the state space, where an atom of the partition corresponds to a signal. This allows one to translate the model into terms of cooperative games: states of nature take the role of players, and atoms of the partition take the role of coalitions. Furthermore, the worth of a coalition is the maximal utility achievable on the corresponding atom.

When the information structure is stochastic, the translation to cooperative games is not possible anymore. Instead, an information function is expressed by means of another function defined over posteriors. It turns out that the key characteristics of an information function is that this function (defined over the set of posteriors) is *convex* and *continuous*.

Beyond the analysis of the cardinal problem (characterizing value of information functions) in order to study the ordinal issue (binary relation over structures), we resort to a classical theorem of von Neumann and Morgenstern (1947). Von Neumann and Morgenstern (vN-M) characterized the binary relations defined over convex sets that can be represented by affine functions. It turns out that all (vN-M) axioms go through, except for continuity which does not suffice in its original form. Instead, continuity in a stronger sense is needed; the order should be continuous in the weak-* topology.

This paper is organized as follows. In Section 2 the model, the ordinal order and the cardinal order over information structures are presented. Section 3 contains the main results: characterization of the binary orders that arise from decision problems and of information functions. Section 4 elaborates on additive separability which is an important property of information functions. The proofs of the main results appear in Section 5. The paper ends with final remarks.

2. THE MODEL

Let $\Omega = \{\omega_1, \dots, \omega_n\}$ be a finite set of states of nature, and let μ be the prior probability over Ω . It is assumed that μ assigns a positive probability to any state. That is, $\mu(\omega) > 0$ for every $\omega \in \Omega$. The set of actions available to the DM is denoted by A . The utility of the DM when she takes the action $a \in A$ and when the state of nature is $\omega \in \Omega$ is denoted by $u(a, \omega)$.

An information structure is a pair $I = (S, M)$, where S is the set of signals and M is a collection of distributions on S , one for each state. M

can be thought of as a stochastic matrix with n rows; the i -th row of M (for $1 \leq i \leq n$) is the distribution over signals given the state ω_i . Stated differently, the cell M_{is} of the matrix M is the probability of receiving the signal $s \in S$ given that the state of nature is ω_i . Note that the number of columns in M coincides with the number of signals in S . For the sake of simplicity, we always write m instead of $|S|$, where no confusion can arise.

Denote by \mathcal{I} the set of all information structures.

For a given information structure $I = (S, M)$, denote by $\pi_I = (\pi_I^s)_{s \in S}$ the distribution on S induced by M . That is, $\pi_I^s = \sum_{i=1}^n \mu(\omega_i) M_{is}$ is the probability of observing s . Also, for $s \in S$, let $q_{I,s} = (q_{I,s}^1, \dots, q_{I,s}^n)$ be the distribution on Ω given that the observed signal is s . Formally, $q_{I,s}^i = \mathbb{P}_I(\omega_i | s) = \frac{\mu(\omega_i) M_{is}}{\pi_I^s}$.

A (pure) strategy of the DM is a function $\sigma : S \rightarrow A$ which dictates the action to be chosen after observing each of the signals. If the information structure is $I = (S, M)$ and the DM follows some strategy σ , then her expected utility is

$$E_{I,\sigma} = \sum_{i=1}^n \mu(\omega_i) \sum_{s \in S} M_{is} u(\sigma(s), \omega_i) = \sum_{s \in S} \pi_I^s \sum_{i=1}^n q_{I,s}^i u(\sigma(s), \omega_i).$$

$\hat{\sigma}_I$ is an ε -optimal strategy, subject to the information structure I , if for every $s \in S$, $\hat{\sigma}_I(s) = a$ implies that $\sum_{i=1}^n q_{I,s}^i u(a, \omega_i) \geq \sum_{i=1}^n q_{I,s}^i u(b, \omega_i) - \varepsilon$ for every $b \in A$. In other words, a strategy is ε -optimal if after observing a signal the action prescribed by the strategy ensures at least the supremum of the expected achievable utility up to an ε . Note that in general, if we allow arbitrary action set and utility function, then there need not be a 0-optimal strategy.

For a certain action set A and a utility function u , let $v_{A,u}(q_{I,s}) = \sup_{a \in A} \sum_{i=1}^n q_{I,s}^i u(a, \omega_i)$ where $I = (S, M)$ is some information structure and $s \in S$. $v_{A,u}$ is the maximal expected utility that the DM can obtain upon observing s (that is, when the posterior distribution is $q_{I,s}$). As a function on $\Delta(\Omega)$, $v_{A,u}$ is the supremum of a set of linear functions and is therefore continuous. Formally,

Lemma 1. *If, for a decision problem with action set A and utility function u , the function $v_{A,u}$ is finite, then it is continuous on $\Delta(\Omega)$.*

The *value* of an information structure $I = (S, M)$ (for a given action set A and a utility function u) denoted $V_{A,u}(I)$, is the maximal expected utility achievable when signals are received according to it. That is, $V_{A,u}(I) = \sum_{s \in S} \pi_I^s v_{A,u}(q_{I,s})$.

The following notation will be useful in the sequel.

Notation 1. For any information structure $I = (S, M)$ and a subset of signals $T \subseteq S$,

(a) $\underline{I}(T) = (\underline{S}(T), \underline{M}(T))$ is the information structure defined as follows: $\underline{S}(T) = (S \setminus T) \cup \{t\}$. If $s \in S \setminus T$ then $\underline{M}(T)_{is} = M_{is}$, and if $s = t$ then $\underline{M}(T)_{is} = \sum_{s' \in T} M_{is'}$, for every $1 \leq i \leq n$.

(b) $\bar{I}(T) = (\bar{S}(T), \bar{M}(T))$ is the information structure defined as follows: Every signal $t \in T$ is replaced by a set of n signals $S_t = \{t_1, \dots, t_n\}$, so the new set of signals is $\bar{S}(T) = (S \setminus T) \cup (\cup_{t \in T} S_t)$. For $s \in S \setminus T$, $\bar{M}(T)_{is} = M_{is}$ ($1 \leq i \leq n$). For $t_k \in S_t$ (for some $t \in T$), if $i = k$ then $\bar{M}(T)_{it_k} = M_{i,t}$ and if $i \neq k$ then $\bar{M}(T)_{it_k} = 0$.

(c) For two disjoint sets of signals $T_1, T_2 \subseteq S$, $\underline{I}(T_1, T_2) = \underline{\underline{I}}(T_1)(T_2)$.

In words, $\underline{I}(T)$ is the information structure which differs from I only in that the columns corresponding to the signals in T are summed up to form one column. Instead of being informed separately of the signals in T , the DM is informed that one of the signals in T occurred. That is, the signals in T are lumped together. On the other hand, $\bar{I}(T)$ is the information structure which is identical to I on $S \setminus T$, and any column corresponding to some $s \in T$ is replaced by a diagonal $n \times n$ matrix.

2.1. Orders induced by decision problems. Let the state space Ω and the prior μ be fixed. Consider a decision problem characterized by an action set A and a utility function u . The DM may obtain information about the realized state through various information structures. Each information structure I entails a different maximal achievable expected payoff $V_{A,u}(I)$. Therefore, every decision problem induces a binary relation \succeq over the set \mathcal{I} of all information structures in the following way. For every $I, I' \in \mathcal{I}$, $I \succeq I'$ if and only if $V_{A,u}(I) \geq V_{A,u}(I')$. Formally,

Definition 1. A binary relation \succeq over \mathcal{I} is induced by a decision problem if there is an action set A and a utility function $u : A \times \Omega \rightarrow \mathbb{R}$ such that $I \succeq I'$ if and only if $V_{A,u}(I) \geq V_{A,u}(I')$.

Our first goal is to characterize the binary relations over \mathcal{I} that are induced by decision problems. These are the relations that can be rationalized by a utility maximization behavior of a Bayesian decision maker.

2.2. Information functions. The second issue considered in this paper is the value of information structures. An outside observer collects data about the value of each information structure for the DM. He cannot observe the DM's prior distribution nor her utility function, while he can observe her payoffs. For instance, this is the case when the only available data is how much an agent is willing to pay in order to use each information structure.

Definition 2. A function $V : \mathcal{I} \rightarrow \mathbb{R}$ is an information function if there exist a set of actions A and a utility function $u : A \times \Omega \rightarrow \mathbb{R}$ such that $V(I) = V_{A,u}(I)$ for every $I \in \mathcal{I}$.

Unlike the ordinal preferences in the previous subsection, here it is assumed that the outside observer knows the exact value of each information structure. Our objective is to find conditions that this (cardinal) data should satisfy in order to represent the worth of information structures in a Bayesian decision problem.

3. THE MAIN RESULTS

3.1. Orders induced by decision problems. In this section we introduce necessary and sufficient conditions for a binary relation over \mathcal{I} to be induced by a decision problem. Let \succeq be a binary relation defined on \mathcal{I} , and let \succ (\sim) be its asymmetric (symmetric) parts.

We will use Kreps' (1988) terminology and will call probability distribution with finite support *simple*. Notice that any prior distribution and an information structure induce a simple probability distribution over $\Delta(\Omega)$. Indeed, for every $\lambda \in \Delta(\Omega)$ and for every $I = (S, M) \in \mathcal{I}$, let $D_I(\lambda) = \{s \in S; q_{I,s} = \lambda\}$. For a fixed information structure I , $D_I(\lambda)$ is not empty only for a finite number of λ 's. The simple probability distribution β_I over $\Delta(\Omega)$ induced by I is $\beta_I(\lambda) = \sum_{s \in D_I(\lambda)} \pi_I^s$ for every $\lambda \in \Delta(\Omega)$.

Denote $\mathcal{B} = \{\beta_I; I \in \mathcal{I}\}$. Notice that not all simple probability distributions over $\Delta(\Omega)$ are in \mathcal{B} . For instance, if $\lambda \in \Delta(\Omega)$, $\lambda \neq \mu$, where μ is the prior of the DM, then the distribution which assigns probability 1 to λ is not in \mathcal{B} (in fact, any $\beta \in \mathcal{B}$ has as expectation the original prior μ). Nevertheless, it is not hard to see that \mathcal{B} is a convex set.

Assume that \succeq is induced by the decision problem with action set A and utility function u . That is $I \succeq I'$ iff $V_{A,u}(I) \geq V_{A,u}(I')$. Recall that $V_{A,u}(I) = \sum_{s \in S} \pi_I^s v_{A,u}(q_{I,s})$. This means that $V_{A,u}(I)$ is the expected value of the function $v_{A,u} : \Delta(\Omega) \rightarrow \mathbb{R}$, where the expectation is taken according to the distribution β_I . Therefore, we can think of $V_{A,u}$ as a function on \mathcal{B} rather than on \mathcal{I} .

The above discussion implies that a necessary condition that \succeq should satisfy in order to be induced by a decision problem is that it depends solely on the distribution over the posteriors. Formally,

Condition 1. (*Reducibility*) If $\beta_I = \beta_{I'}$ then $I \sim I'$.

Next, notice that $V_{A,u}$ (as a function on \mathcal{B}) is affine. That is, $V_{A,u}(\alpha\beta + (1 - \alpha)\beta') = \alpha V_{A,u}(\beta) + (1 - \alpha)V_{A,u}(\beta')$ for every $\beta, \beta' \in \mathcal{B}$ and for every $0 \leq \alpha \leq 1$. The von Neumann-Morgenstern theorem (von Neumann and Morgenstern, 1947) provides necessary and sufficient conditions for a binary

relation over a convex set to be represented by an affine real-valued function. Note that, by Condition 1, we can think of \succeq as a binary relation over the convex set \mathcal{B} . It is important to note, once again, that unlike the vast majority of the literature (see e.g., Kreps (1988) Corollary 5.12), here \succeq is not defined over all simple distributions but over a subset of them. As a result, the proof technique is also different.

We state appropriate versions of the von Neumann-Morgenstern axioms. The first two conditions (Weak Order and Independence) are standard.

Condition 2. (*Weak Order*) \succeq is a weak order over \mathcal{B} .

Condition 3. (*Independence*) If $\beta \succ \beta'$ then $\alpha\beta + (1-\alpha)\beta'' \succ \alpha\beta' + (1-\alpha)\beta''$ for all $\beta'' \in \mathcal{B}$ and for all $\alpha \in (0, 1)$.

The following condition states that the order \succeq is continuous. That is, if $\beta \succ \beta'$, then for any β'' sufficiently close to β , $\beta'' \succ \beta'$, and moreover, for any β'' sufficiently close to β' , $\beta \succ \beta''$. The question is what would be the appropriate meaning of “sufficiently close to”? It turns out that the weak-* topology is the right notion.

Roughly speaking, the weak-* topology refers to the simple distributions in \mathcal{B} as measures with respect to which continuous functions are integrated. Two simple distributions are considered close if the integrals of continuous functions with respect to both are close. Formally, the weak-* topology is defined by a basis of open sets. A typical open set in the basis is determined by three parameters: $\beta \in \mathcal{B}$, $\varepsilon > 0$ and a continuous function f defined over $\Delta(\Omega)$, and is defined as $O(\beta, \varepsilon, f) = \{\beta' \in \mathcal{B}; |\int f d\beta - \int f d\beta'| < \varepsilon\}$. An open set in the weak-* topology is therefore a finite intersection of sets of the kind $O(\beta, \varepsilon, f)$.

Condition 4. (*Continuity*) For any β the sets $\{\beta'; \beta \succ \beta'\}$ and $\{\beta'; \beta' \succ \beta\}$ are open in the weak-* topology.

It turns out that another condition is needed for the characterization. Namely, \succeq should be convex in a certain sense. For the statement of this last condition it is convenient to rethink \succeq as a binary relation over \mathcal{I} (recall Notation 1).

Condition 5. (*Convexity*) For every $I = (S, M) \in \mathcal{I}$ and for every $T \subseteq S$, $I \succeq \underline{I}(T)$.

The choice of the term convexity will become clear later on. To understand why this is a necessary condition, recall the equivalence result of Blackwell (1951, 1953). Blackwell defined the *more informative* partial order over information structures. Let $I = (S, M)$ and $I' = (S', M')$ be two information structures. I is *more informative* than I' if there is a stochastic

matrix,³ say C , such that $M' = MC$. That is, M' can be obtained by multiplying M with a stochastic matrix. Blackwell (1951, 1953) showed that I is more informative than I' iff for any decision problem with action set A and utility function u , $V_{A,u}(I) \geq V_{A,u}(I')$. Notice that I is more informative than $\underline{I}(T)$. It follows that if \succeq is induced by a decision problem then it must be convex.

Theorem 1. *A binary relation \succeq over \mathcal{I} satisfies Conditions 1-5 if and only if it is induced by a decision problem.*

The proof of this theorem will follow from the next theorem together with the vN-M theorem (von Neumann-Morgenstern, 1947).

3.2. Characterizing information functions. Here we assume that an outside observer knows the exact value of each information structure for the DM. We characterize those functions defined on information structures that arise from decision problems. The first condition is convexity which is identical to the one in the previous subsection.

Definition 3. *A function $V : \mathcal{I} \rightarrow \mathbb{R}$ is convex if $V(\underline{I}(T)) \leq V(I)$ for every information structure $I = (S, M)$ and for every subset of signals $T \subseteq S$.*

Next, continuity of V is required. Here, the condition is slightly different than in the previous subsection.

Definition 4. *$V : \mathcal{I} \rightarrow \mathbb{R}$ is continuous if for every fixed set of signals S , V is a continuous function of the stochastic matrix M .*

The last condition needed is Additive Separability. It states that $V(I)$ is the expected value of some other function defined on the set of posteriors. Formally,

Definition 5. *$V : \mathcal{I} \rightarrow \mathbb{R}$ is additively separable if there exist a function $v : \Delta(\Omega) \rightarrow \mathbb{R}$ such that $V(I) = \sum_{s \in S} \pi_I^s v(q_{I,s})$ for every $I \in \mathcal{I}$. If v is such a function we will say that v corresponds to V .*

We are now ready to characterize information functions. The proof of the following theorem is postponed to Section 5.

Theorem 2. *$V : \mathcal{I} \rightarrow \mathbb{R}$ is an information function if and only if it is additively separable, convex and continuous.*

³A stochastic matrix is a matrix whose entries are all non-negative and the sum of each row is 1.

4. MORE ON ADDITIVE SEPARABILITY

It is obvious that any information function is additively separable. However, this condition is hard to interpret. We would like to find natural conditions on a function $V : \mathcal{I} \rightarrow \mathbb{R}$ that are equivalent to additive separability. For the next definition recall Notation 1.

Definition 6. A function $V : \mathcal{I} \rightarrow \mathbb{R}$ is Independent of Irrelevant Signals (IIS) if

$$(1) \quad V(\underline{I}(T_1)) + V(\underline{I}(T_2)) = V(\underline{I}(T_1, T_2)) + V(I)$$

for every information structure $I = (S, M)$ and for every two disjoint subsets of signals $T_1, T_2 \subseteq S$.

To justify the term *IIS*, notice that equation (1) can be rewritten as $V(I) - V(\underline{I}(T_1)) = V(\underline{I}(T_2)) - V(\underline{I}(T_1, T_2))$. The left-hand side of this equation is equal to the loss incurred to the DM due to coarsening the information structure: instead of being informed of each signal in T_1 separately, the signals of T_1 are lumped together. This is the value of the information embedded in the set T_1 when the information structure is I . If equation (1) holds for every information structure I and for every subset of signals T_2 , it means that this value is independent of I . This is so, because when $T_2 = S \setminus T_1$, the right-hand side of equation (1) depends only on T_1 . Therefore, the left-hand side is constant across all information structures that contain T_1 . This implies that the contribution of a set of signals (columns in the stochastic matrix) to the value of information is independent of the informational structure out of this set.

Next, recall the reducibility condition of subsection 3.1. We say that a function V is reducible if the order it induces satisfies the reducibility condition. Formally,

Definition 7. A function $V : \mathcal{I} \rightarrow \mathbb{R}$ is reducible if $\beta_I = \beta_{I'}$ implies $V(I) = V(I')$ for every $I, I' \in \mathcal{I}$.

Proposition 1. $V : \mathcal{I} \rightarrow \mathbb{R}$ is additively separable iff it is IIS and reducible.

Proof. Assume first that V is additively separable. Then there exists $v : \Delta(\Omega) \rightarrow \mathbb{R}$ such that $V(I) = \sum_{s \in S} \pi_I^s v(q_{I,s})$ for every $I \in \mathcal{I}$. It is straightforward to see that V is reducible. To check that V is IIS, fix some $I \in \mathcal{I}$

and let T_1, T_2 be two disjoint subsets of signals. We have,

$$\begin{aligned}
V(\underline{I}(T_1)) + V(\underline{I}(T_2)) &= \sum_{s \in \underline{I}(T_1)} \pi_{\underline{I}(T_1)}^s v(q_{\underline{I}(T_1),s}) + \sum_{s \in \underline{I}(T_2)} \pi_{\underline{I}(T_2)}^s v(q_{\underline{I}(T_2),s}) = \\
&= \sum_{s \notin T_1} \pi_I^s v(q_{I,s}) + \pi_{\underline{I}(T_1)}^{t_1} v(q_{\underline{I}(T_1),t_1}) + \sum_{s \notin T_2} \pi_I^s v(q_{I,s}) + \pi_{\underline{I}(T_2)}^{t_2} v(q_{\underline{I}(T_2),t_2}) = \\
&= V(I) + \sum_{s \notin T_1 \cup T_2} \pi_I^s v(q_{I,s}) + \pi_{\underline{I}(T_1)}^{t_1} v(q_{\underline{I}(T_1),t_1}) + \pi_{\underline{I}(T_2)}^{t_2} v(q_{\underline{I}(T_2),t_2}) = \\
&= V(I) + V(\underline{I}(T_1, T_2)).
\end{aligned}$$

Conversely, assume that V is IIS and reducible. We need to prove the existence of a function $v : \Delta(\Omega) \rightarrow \mathbb{R}$ such that $V(I) = \sum_{s \in S} \pi_I^s v(q_{I,s})$ for every $I \in \mathcal{I}$.

In order to find an appropriate function v , we first need to define two auxiliary information structures for every vector $x = (x_1, \dots, x_n)$ with $0 \leq x_i \leq 1$, $i = 1, 2, \dots, n$. The first one is denoted $B^{x,1}$ and has $n+1$ signals. The stochastic matrix is

$$\begin{pmatrix} x_1 & 0 & 0 & \dots & 1-x_1 \\ 0 & x_2 & 0 & \dots & 1-x_2 \\ \vdots & \vdots & \ddots & & \vdots \\ 0 & 0 & 0 & x_n & 1-x_n \end{pmatrix}$$

The second information structure is denoted $B^{x,2}$ and has only 2 signals. The distribution over signals is defined by the matrix

$$\begin{pmatrix} x_1 & 1-x_1 \\ x_2 & 1-x_2 \\ \vdots & \vdots \\ x_n & 1-x_n \end{pmatrix}$$

Finally, let I_d denote the deterministic information structure with n signals, under which the DM is fully informed about the true state of nature.

Notice that, since the prior distribution on Ω is μ , $q_{I,s}$ is always of the form⁴ $q_{I,s} = \frac{\mu \circ x}{\|\mu \circ x\|_1}$, where x is a vector as above. We define $v\left(\frac{\mu \circ x}{\|\mu \circ x\|_1}\right) = V(I_d) - \frac{V(B^{x,1}) - V(B^{x,2})}{\|\mu \circ x\|_1}$. To finish the proof, it only remains to check that if v is defined as above then $V(I) = \sum_{s \in S} \pi_I^s v(q_{I,s})$ for every $I \in \mathcal{I}$. Indeed,

⁴For any two vectors $x = (x_1, \dots, x_n)$ and $y = (y_1, \dots, y_n)$, $x \circ y$ denotes the vector of the same length whose i -th coordinate is equal to $x_i y_i$.

using IIS we have for any $I = (S, M)$ (recall Notation 1(b)),

$$\begin{aligned}
\sum_{s \in S} \pi_I^s v(q_{I,s}) &= \sum_{s \in S} \pi_I^s \left(V(I_d) - \frac{V(B^{M \cdot s, 1}) - V(B^{M \cdot s, 2})}{\|\mu \circ M \cdot s\|_1} \right) = \\
&= V(I_d) - \sum_{s \in S} (V(B^{M \cdot s, 1}) - V(B^{M \cdot s, 2})) = \\
&= V(I_d) - \sum_{s \in S} (V(\bar{I}(\{s\})) - V(I)).
\end{aligned}$$

Recursive use of IIS gives $\sum_{s \in S} (V(\bar{I}(\{s\})) - V(I)) = V(\bar{I}(S)) - V(I)$. Therefore, $\sum_{s \in S} \pi_I^s v(q_{I,s}) = V(I_d) - V(\bar{I}(S)) + V(I)$. It only remains to check that $V(I_d) = V(\bar{I}(S))$. However, since V is reducible we are done. ■

Proposition 1 and Theorem 2 lead to the following conclusion.

Corollary 1. *$V : \mathcal{I} \rightarrow \mathbb{R}$ is an information function iff it is convex, continuous, reducible and IIS.*

We conclude this section with a short discussion on the uniqueness of the function v . A function $v : \Delta(\Omega) \rightarrow \mathbb{R}$ uniquely determines a function $V : \mathcal{I} \rightarrow \mathbb{R}$ via the equation $V(I) = \sum_{s \in S} \pi_I^s v(q_{I,s})$. However, given some additively separable function V , the corresponding v is not unique. The following proposition states that v_1 and v_2 both correspond to V iff $v_1 - v_2$ is a linear function which vanishes at the prior distribution μ .

Proposition 2. *Assume that $V : \mathcal{I} \rightarrow \mathbb{R}$ is additively separable with v_1 corresponding to it. Then, v_2 also corresponds to V , if and only if there exists a vector $x \in \mathbb{R}^n$ such that⁵ $x\mu = 0$ and $v_1(q) - v_2(q) = xq$ for every $q \in \Delta(\Omega)$.*

Proof. Assume that for a certain $x \in \mathbb{R}^n$ $v_2(q) = v_1(q) - xq$, for every $q \in \Delta(\Omega)$. Moreover, assume that $x\mu = 0$. We show that v_2 corresponds to V . For every $I \in \mathcal{I}$,

$$\begin{aligned}
\sum_{s \in S} \pi_I^s v_2(q_{I,s}) &= \sum_{s \in S} \pi_I^s (v_1(q_{I,s}) - xq_{I,s}) = V(I) - \sum_{s \in S} \pi_I^s xq_{I,s} = \\
&= V(I) - \sum_{s \in S} \pi_I^s \sum_{i=1}^n x^i \frac{\mu(\omega_i) M_{is}}{\pi_I^s} = \\
&= V(I) - \sum_{i=1}^n x^i \mu(\omega_i) \sum_{s \in S} M_{is} = V(I) - x\mu = V(I).
\end{aligned}$$

In the other direction, assume that v_2 corresponds to V and define $v = v_1 - v_2$. Let $q_j = (q_j^1, \dots, q_j^n)$, $j = 1, 2$, be two distributions over Ω , and

⁵For any two vectors $x = (x_1, \dots, x_n)$ and $y = (y_1, \dots, y_n)$, xy denotes the inner product $\sum x_i y_i$.

let $\alpha \in [0, 1]$. For $i = 1, \dots, n$, define $r_1^i = c\alpha \frac{q_1^i}{\mu(\omega_i)}$ and $r_2^i = c(1 - \alpha) \frac{q_2^i}{\mu(\omega_i)}$, where c is a positive constant that satisfies $r_1^i + r_2^i \leq 1$ for every $i = 1, \dots, n$. Finally, define $r_3 = (1, \dots, 1) - r_1 - r_2$. Consider the information structure $I = (S, M)$, where $S = \{s_1, s_2, s_3\}$, and M is the $n \times 3$ matrix whose j 'th column is r_j , $j = 1, 2, 3$. Since both, v_1 and v_2 , correspond to V , we have for $j = 1, 2$, $V(I) = \sum_{s \in S} \pi_I^s v_j(q_{I,s}) = c\alpha v_j(q_1) + c(1 - \alpha)v_j(q_2) + (1 - c)v_j(q_3)$, where $q_3 = \frac{r_3 \circ \mu}{\|r_3 \circ \mu\|_1}$. Reorganizing the terms yields,

$$(2) \quad c\alpha v(q_1) + c(1 - \alpha)v(q_2) = (c - 1)v(q_3).$$

Set $T = \{s_1, s_2\}$. For $j = 1, 2$ we obtain, $V(I(T)) = cv_j(\alpha q_1 + (1 - \alpha)q_2) + (1 - c)v_j(q_3)$, which is equivalent to

$$(3) \quad cv(\alpha q_1 + (1 - \alpha)q_2) = (c - 1)v(q_3).$$

From (2) and (3) it follows that $\alpha v(q_1) + (1 - \alpha)v(q_2) = v(\alpha q_1 + (1 - \alpha)q_2)$ for every two distributions q_1, q_2 and for every $\alpha \in [0, 1]$. In other words, $v = v_1 - v_2$ is linear on $\Delta(\Omega)$. Thus, there is $x \in \mathbb{R}^n$ such that $v_1(q) - v_2(q) = xq$ (Since v is only defined on the simplex, we can always extend it to a linear function which passes through the origin). Finally, since v_1 and v_2 agree on the information structure with only one signal, $x\mu = v_1(\mu) - v_2(\mu) = 0$. ■

The result of Proposition 2 might be interpreted in the following way. Given the prior μ and an information structure I , the DM forms a probability distribution over the set of posteriors (this is the distribution β_I of subsection 3.1). It is easy to see that the expectation of this probability distribution over posteriors is the prior μ . Now, if $v_2(q) = v_1(q) + xq$, where $x\mu = 0$, then an expected utility maximizer is indifferent between v_1 and v_2 . Indeed, the expected contribution of the difference xq is always 0 since $\sum_{s \in S} \pi_I^s xq_{I,s} = x\mu = 0$.

5. THE PROOFS OF THE THEOREMS

Before proving the main theorems, we first need several lemmas. The first lemma explains the choice of the name convexity in Definition 3.

Lemma 2. *Let $V : \mathcal{I} \rightarrow \mathbb{R}$ be additively separable function and v corresponds to V . Then, V is convex if and only if v is convex on $\Delta(\Omega)$.*

Proof. Let $q_j = (q_j^1, \dots, q_j^n)$, $j = 1, 2$, be two distributions over Ω , and let $\alpha \in [0, 1]$. We start by showing that if V is convex then $\alpha v(q_1) + (1 - \alpha)v(q_2) \geq v(\alpha q_1 + (1 - \alpha)q_2)$.

For $i = 1, \dots, n$, define $r_1^i = c\alpha \frac{q_1^i}{\mu(\omega_i)}$ and $r_2^i = c(1 - \alpha) \frac{q_2^i}{\mu(\omega_i)}$, where c is a positive constant that satisfies $r_1^i + r_2^i \leq 1$ for every $i = 1, \dots, n$. Let $r_3 = (1, \dots, 1) - r_1 - r_2$, and consider the information structure $I = (S, M)$

where $S = \{s_1, s_2, s_3\}$ and M is an $n \times 3$ matrix whose j -th column is r_j , $j = 1, 2, 3$.

Note that $V(I) = \sum_{s \in S} \pi_I^s v(q_{I,s}) = c\alpha v(q_1) + c(1-\alpha)v(q_2) + (1-c)v(q_3)$, where $q_3 = \frac{r_3 \circ \mu}{\|r_3 \circ \mu\|_1}$. If $T = \{s_1, s_2\}$ then $V(\underline{I}(T)) = cv(\alpha q_1 + (1-\alpha)q_2) + (1-c)v(q_3)$. Since V is convex, $V(I) \geq V(\underline{I}(T))$. Thus, $c\alpha v(q_1) + c(1-\alpha)v(q_2) + (1-c)v(q_3) \geq cv(\alpha q_1 + (1-\alpha)q_2) + (1-c)v(q_3)$, which implies that $\alpha v(q_1) + (1-\alpha)v(q_2) \geq v(\alpha q_1 + (1-\alpha)q_2)$.

In the other direction, assume that v is convex and let $T \subseteq S$ be a subset of signals of some information structure $I = (S, M)$. Recall that the set of signals in the information structure $\underline{I}(T)$ is $(S \setminus T) \cup \{t\}$, and the column corresponding to the signal t is the sum of columns of the signals in T . By convexity of v we obtain

$$\begin{aligned} V(\underline{I}(T)) &= \sum_{s \in S \setminus T} \pi_I^s v(q_{I,s}) + \pi^t v(q_t) \\ &\leq \sum_{s \in S \setminus T} \pi_I^s v(q_{I,s}) + \pi^t \sum_{s \in T} \frac{\pi_I^s}{\pi^t} v(q_{I,s}) = V(I). \end{aligned}$$

Therefore, V is convex. ■

The following lemma links between the continuity of V and any v that corresponds to it.

Lemma 3. *Let $V : \mathcal{I} \rightarrow \mathbb{R}$ and let v correspond to V . Then V is continuous iff v is continuous on $\Delta(\Omega)$.*

Proof. If v is continuous on $\Delta(\Omega)$ then, since both π_I^s and $q_{I,s}$ are continuous functions of the matrix M , V is also continuous. On the other hand, assume that V is continuous and let $q_1, q_2 \in \Delta(\Omega)$ be close. Consider the information structures $I_j = (S_j, M_j)$, $j = 1, 2$, where the first column of the matrix M_j is the vector $r_j^i = c \frac{q_j^i}{\mu(\omega_i)}$, $1 \leq i \leq n$ (c is a positive constant such that $r_j^i \leq 1$ for all $1 \leq i \leq n$, $j = 1, 2$), and the rest of the columns are such that the DM is fully informed on the true state of nature. We have⁶ $V(I_j) = cv(q_j) + \sum_{i=1}^n (\mu(\omega_i) - cq_j^i) v(\mathbb{1}_{\omega_i})$, $j = 1, 2$. Thus, $|V(I_1) - V(I_2)| = |c(v(q_1) - v(q_2)) + \sum_{i=1}^n c(q_1^i - q_2^i) v(\mathbb{1}_{\omega_i})|$. If q_1, q_2 are close, then the stochastic matrices M_1, M_2 are close, and by assumption $|V(I_1) - V(I_2)|$ is small. Since $|\sum_{i=1}^n c(q_1^i - q_2^i) v(\mathbb{1}_{\omega_i})|$ is small when q_1, q_2 are close, it must be that $|v(q_1) - v(q_2)|$ is also small, so v is continuous. ■

We are now ready to prove the main results. We start with the cardinal characterization (Theorem 2) which will be then used to prove Theorem 1.

⁶ $\mathbb{1}_{\omega_i}$ is the probability distribution over Ω which assigns probability 1 to ω_i .

Proof of Theorem 2. Assume first that V is an information function. Obviously it is additively separable. Since every information function is monotonic with respect to the ‘more informative’ order, V is convex. Finally, since V is a real-valued function, it follows that the corresponding function v is finite on $\Delta(\Omega)$. By Lemma 1, v is continuous and by Lemma 3, V is continuous.

Conversely, assume that V is additively separable, convex and continuous function, and let v correspond to V . By Lemmas 2 and 3, v is convex and continuous on $\Delta(\Omega)$. Therefore, at every point r in the interior of the simplex, there is a vector $x_r = (x_r^1, \dots, x_r^n)$ (that defines the tangent to the graph of v at the point r) such that $v(q) \geq qx_r$ for every $q \in \Delta(\Omega)$ with equality when $q = r$. In particular, $v(q) = \sup_r qx_r$ for every $q \in \Delta(\Omega)$ (note that, by continuity, the last equality holds also for points q in the boundary of the simplex).

Define the set of actions, A , to be identical to the interior of the simplex. When the state realized is ω_i and the action taken is r , the utility, $u(r, \omega_i)$, is defined to be x_r^i . Thus, when the distribution over states is $q = (q^1, \dots, q^n)$ and the action taken is r , the expected utility is $\sum_i q^i u(r, \omega_i) = qx_r$. Hence, when the posterior distribution over states is q , the maximal expected utility achievable is exactly $v(q)$. ■

Proof of Theorem 1. Continuity implies the Archimedian condition of von Neumann and Morgenstern which states that if $\beta \succ \beta' \succ \beta''$, then there are $\alpha, \gamma \in (0, 1)$ such that $\alpha\beta + (1 - \alpha)\beta'' \succ \beta' \succ \gamma\beta + (1 - \gamma)\beta''$. Indeed, let $\beta \succ \beta' \succ \beta''$. By continuity, the set $\{\beta; \beta \succ \beta'\}$ is open in the weak-* topology. In particular, there are finitely many basis open sets $O(\beta_i, \varepsilon_i, f_i)$, $i = 1, \dots, \ell$, such that $\beta \in \cap_{i=1}^\ell O(\beta_i, \varepsilon_i, f_i)$ and, moreover, every $\tilde{\beta} \in \cap_{i=1}^\ell O(\beta_i, \varepsilon_i, f_i)$ satisfies $\tilde{\beta} \succ \beta'$.

For every $i = 1, \dots, \ell$ there is a sufficiently large $\alpha_i < 1$ such that if $\alpha > \alpha_i$, then $|\int f_i d(\alpha\beta + (1 - \alpha)\beta'') - \int f_i d\beta| < \varepsilon_i - |\int f_i d\beta_i - \int f_i d\beta|$. Thus, $|\int f_i d(\alpha\beta + (1 - \alpha)\beta'') - \int f_i d\beta_i| < \varepsilon_i$. This implies that $\alpha\beta + (1 - \alpha)\beta'' \in O(\beta_i, \varepsilon_i, f_i)$. Hence, if $\alpha > \max_{1 \leq i \leq \ell} \{\alpha_i\}$, then $\alpha\beta + (1 - \alpha)\beta'' \in \cap_{i=1}^\ell O(\beta_i, \varepsilon_i, f_i)$, which implies that $\alpha\beta + (1 - \alpha)\beta'' \succ \beta'$, as desired. By employing a similar argument, one can find γ such that $\beta' \succ \gamma\beta + (1 - \gamma)\beta''$.

The von Neumann-Morgenstern theorem (see Theorem 5.11 in Kreps, 1988) implies that the Weak order, Archimedian and Independence conditions ensure that \succeq is represented by an affine function V . In order to prove that V is a value of information function, by Theorem 2, it is sufficient to show that it is additively separable, convex and continuous.

In order to prove that V is additively separable we use Proposition 1 and show that V is Independent of Irrelevant Signals (recall Definition 6). For this purpose fix an information structure $I = (S, M)$ and two disjoint

subsets of signals $T_1, T_2 \subseteq S$. Since, $\frac{1}{2}\beta_{\underline{I}(T_1)} + \frac{1}{2}\beta_{\underline{I}(T_2)} = \frac{1}{2}\beta_{\underline{I}(T_1, T_2)} + \frac{1}{2}\beta_I$, affinity of V guarantees that $V(\underline{I}(T_1)) + V(\underline{I}(T_2)) = V(\underline{I}(T_1, T_2)) + V(I)$. Thus, V is Independent of Irrelevant Signals.

Due to condition 5, $I \succeq \underline{I}(T)$. Thus, $V(I) \geq V(\underline{I}(T))$ and therefore, V is convex. Finally, we need to show continuity of V (recall Definition 4). Fix a set of signals S and let $I_n = (S, M_n)$ be a sequence of information structures. Assume that $M_n \rightarrow M$. We show that $V(\beta_{I_n}) \rightarrow V(\beta_I)$. Otherwise, either $\limsup V(\beta_{I_n}) > V(\beta_I)$ or $\liminf V(\beta_{I_n}) < V(\beta_I)$. If $\limsup V(\beta_{I_n}) > V(\beta_I)$, then, by affinity of V , there is an m sufficiently large so that $\limsup V(\beta_{I_n}) > V(\frac{1}{2}\beta_{I_m} + \frac{1}{2}\beta_I) > V(\beta_I)$. Thus, the set $O = \{\beta \mid \frac{1}{2}\beta_{I_m} + \frac{1}{2}\beta_I \succ \beta\}$ contains β_I and, by assumption, is open.

Since $M_n \rightarrow M$, for sufficiently large n , $\beta_{I_n} \in O$. Thus, $V(\frac{1}{2}\beta_{I_m} + \frac{1}{2}\beta_I) > V(\beta_{I_n})$. This implies that $V(\frac{1}{2}\beta_{I_m} + \frac{1}{2}\beta_I) \geq \limsup V(\beta_{I_n})$, which contradicts the choice of m . It implies that $\limsup V(\beta_{I_n}) \leq V(\beta_I)$. By a similar method one can show that $\liminf V(\beta_{I_n}) \geq V(\beta_I)$, and hence, $V(\beta_{I_n}) \rightarrow V(\beta_I)$ which proves that V is continuous.

We therefore conclude that V is additively separable, convex and continuous and is therefore an information function. ■

6. FINAL REMARKS

6.1. Information functions with *some* prior. Theorem 2 refers to the case where the state space Ω and the prior distribution μ are known to the outside observer. It characterizes those observations that are consistent with a rational behavior of a decision maker in a Bayesian setting, given that the prior is μ . One may ask a similar question for an unknown prior. That is, when are the observations consistent with a rational behavior of a decision maker with *some* prior μ ?

Regarding Ω , if we assume that the outside observer can see the information structure, then we implicitly assume that he can also see the state space (or at least its cardinality). However, the observer need not know the prior beliefs of the DM.

The properties *IIS*, *convexity* and *continuity* of an information function V do not depend on the prior distribution μ . The *reducibility* condition, although phrased in terms of the distributions β_I (which depends on the prior), could be rephrased without resorting to any particular prior: for any pair of signals $s_1, s_2 \in S$ whose corresponding columns are proportional, $V(\underline{I}(\{s_1, s_2\})) = V(I)$. This implies that if V has these four properties, then for any distribution μ , as long as it has a full support (all the states are assigned positive probability), V is an information function of a decision problem with μ being its prior. In other words, V is an information

function with *a certain* prior having full support if and only if it is an information function with *any* prior having full support. Furthermore, if V is an information function with *a certain* prior (not necessarily with a full support), then by assigning utility 0 to any 0-probability state, one may obtain a representation of V as an information function with a prior having a full support. It should be noted, however, that the corresponding v might change with the prior.

6.2. Non-uniqueness of the utility function. The utility function used in the proof of Theorem 2 is defined by the tangent hyperplanes to the graph of the function v that corresponds to V . However, by Proposition 2 the corresponding v is not unique. More precisely, two functions $v_1, v_2 : \Delta(\Omega) \rightarrow \mathbb{R}$ correspond to the same information function V iff the difference between them is a linear function which vanishes at the prior distribution μ . It follows that, for a given information function V , the utility function $u : A \times \Omega \rightarrow \mathbb{R}$ which induces V is not unique. Indeed, if $x \in \mathbb{R}^n$ satisfies $x\mu = 0$ then the utility function \tilde{u} defined by $\tilde{u}(a, \omega_i) = u(a, \omega_i) + x^i$ will result in the same information function as the one induced by u .

When the data is ordinal this non-uniqueness becomes even more drastic. Indeed, the proof of Theorem 1 uses the vN-M theorem in order to get the representation of \succeq by an affine function V on the set \mathcal{B} . It is known that \tilde{V} also represents \succeq iff there are $\alpha > 0$ and $K \in \mathbb{R}$ such that $\tilde{V} = \alpha V + K$. It follows that, if $v : \Delta(\Omega) \rightarrow \mathbb{R}$ represents \succeq , then \tilde{v} also represents \succeq iff $\tilde{v}(q) = \alpha v(q) + xq$ for some $\alpha > 0$ and $x \in \mathbb{R}^n$ (not necessarily with $x\mu = 0$). This implies that similar transformations of the utility function u (multiplying the entire function by some positive constant and adding the same vector to the utility of every action) will not affect the induced order.

6.3. Alternative characterizations. By Lemmas 2 and 3, Theorem 2 can be rephrased as follows: *$V : \mathcal{I} \rightarrow \mathbb{R}$ is an information function if and only if it is additively separable and every function v that corresponds to it is convex and continuous.*

Another possible statement of the result in Theorem 2 is the following: *$V : \mathcal{I} \rightarrow \mathbb{R}$ is an information function if and only if it is additively separable, continuous and monotonic with respect to the ‘more informative’ order.*

6.4. Convexity and Monotonicity. As explained in the previous sections (see the discussion after Condition 5 of subsection 3.1), if a function $V : \mathcal{I} \rightarrow \mathbb{R}$ is monotonically non-decreasing with respect to Blackwell’s ‘more informative’ order then it is convex. By Theorem 2, under the additional conditions of additive separability and continuity, the converse also holds. That is, if an additively separable and continuous function V is convex then

it is monotonically non-decreasing with respect to the ‘more informative’ order. It seems that a direct proof of this point is not easy.

6.5. The role of continuity. To illustrate the role of continuity, consider the (convex) function $v : \Delta(\Omega) \rightarrow \mathbb{R}$ which equals 1 on the vertices of the simplex and 0 elsewhere. The induced function V assigns to every information structure the probability that the DM will know for sure the true state of nature under this structure. Although V is additively separable and convex it is not an information function. This follows from Lemma 1.

6.6. Compact action set and continuous utility function. The action set A defined in the proof of Theorem 2 is the interior of the simplex of distributions over Ω , while the utility function u is defined by the tangent hyperplanes to the convex function v . Thus, the action set is not compact and the utility function is not necessarily continuous (however, since v is almost surely differentiable, the utility u is almost surely continuous on $A \times \Omega$). The following example shows that, if one commits for a compact action set and a continuous utility function, then the conditions of Theorem 2 may not suffice.

Example 1: The entropy function (Shannon, 1948) is defined by $e(q) = \sum_{i=1}^n -q_i \log(q_i)$ for $q \in \Delta(\Omega)$ (if $q_i = 0$ for some $1 \leq i \leq n$ then $q_i \log(q_i) = 0$). Since e is concave, consider the convex function $-e$. Obviously, the function V induced by $-e$ satisfies the three conditions of Theorem 2 and, therefore, it is an information function. We claim, however, that it cannot be an information function of a decision problem having a compact action set and a continuous utility function. The reason is that the derivative of $-e$ is not bounded near the boundary of the simplex. Thus, one cannot extend the utility in a continuous way to the boundary of the simplex. ■

The above example suggests that in order to extend the utility to the boundary of the simplex in a continuous way, v should satisfy an additional condition. Namely, v is Lipschitz (i.e., there is a constant $K > 0$ such that for every q, q' in the simplex, $|v(q) - v(q')| \leq K\|q - q'\|$, where $\|\cdot\|$ is any norm), as stated (without a proof) in the following proposition.

Theorem 3. *A function $V : \mathcal{I} \rightarrow \mathbb{R}$ is an information function of a decision problem with a compact action set and a continuous utility function if and only if it is additively separable, and any function v corresponding to it is continuous, Lipschitz and convex.*

6.7. Finite action set.

Definition 8. A real function v defined on the simplex is piecewise-linear if there are finitely many disjoint sets W_1, \dots, W_k in the simplex, such that⁷ $\cup_{i=1}^k \text{cl}W_i$ covers the entire simplex and v is linear over W_i , $i = 1, \dots, k$.

The proof of Theorem 2 suggests that a function $V : \mathcal{I} \rightarrow \mathbb{R}$ is an information function of a decision problem with finitely many actions if and only if it is additively separable and any v corresponding to it is a piecewise-linear, continuous and convex function.

6.8. Measuring information by convex functions. The partial order ‘more informative than’ over information structures can be defined in several equivalent ways (see Le Cam, 1996). For a continuous and convex function $w : \Delta(\Omega) \rightarrow \mathbb{R}$, we say that ‘ $I = (S, M)$ contains at least as much information as $I' = (S', M')$ with respect to w ’ if $\sum_{s \in S} \pi_I^s w(q_{I,s}) \geq \sum_{s \in S'} \pi_{I'}^s w(q_{I',s})$. It is known (see Le Cam, Theorem 1, page 130) that I is more informative than I' if and only if I contains at least as much information as I' with respect to w , for every convex and continuous function w .

The above equivalence result can be proved with the help of Theorem 2. Indeed, by Theorem 2, I contains at least as much information as I' with respect to w for every convex and continuous function w , iff for any decision problem, I yields higher expected utility than I' . Thus, since I is better than I' in any decision problem iff I is more informative than I' , the equivalence follows.

7. REFERENCES

- Blackwell, D. (1951) “Comparison of Experiments,” *Proc. Second Berkeley Symposium on Math. Stat and Prob.*, University of California Press, 93-102.
- Blackwell, D. (1953) “Equivalent Comparison of Experiments,” *Ann. Math. Stat.*, **24**, 265-272.
- Gilboa, I. and E. Lehrer (1991) “The value of information – An axiomatic approach,” *J. Math. Economics*, **20**, 443-459.
- Kreps, D. (1988) *Notes on the Theory of Choice*, Westview Press.
- Le Cam, L. (1996) “Comparison of experiments - A short review,” in *Statistics, Probability and Game Theory—Papers in Honor of David Blackwell*, (T. Ferguson, L. Shapley, eds.), IMS Lecture Notes - Monograph Series 30.
- Shannon, C. (1948) “The mathematical theory of communication,” *Bell System Technical Journal*, **27**, 379-423.
- Torgersen, E.N. (1991) *Comparison of Statistical Experiments*, Cambridge Univ. Press.

⁷ $\text{cl}W$ is the closure of W .

von Neumann, J. and O. Morgenstern (1947) “Theory of games and economic behavior”, 2nd edition, Princeton: Princeton University Press.