Information and the Risk-Averse Firm

by

Robert G. Chambers and John Quiggin

WP 01-03

Department of Agricultural and Resource Economics
The University of Maryland, College Park

Information and the Risk-Averse Firm

Robert G. Chambers[1] and John Quiggin[2]

February 26, 2001

[1]Professor of Agricultural Economics, University of Maryland, College Park, and Adjunct Professor Agricultural Economics, University of Western Australia

[2]Australian Research Council Fellow, Australian National University

1 Introduction

Information plays a central role in a modern economy. In any problem involving decisions under uncertainty, the choice of action and the expected benefits of alternative choices depend on the information available to the decision-maker. Economic analysis of information began with the classic work Blackwell (1951), who showed that the value of information is always positive for expected utility preferences.

As with analysis of choice under uncertainty in general, most further discussion of the value of information has focused either on choice from a finite set of alternatives or on cases where the decision variable is a scalar such as effort. An especially important example of the latter case is that of a firm facing either price uncertainty or technological uncertainty represented by a stochastic production function technology. Scalar decision problems for producers facing either price or production uncertainty have been analyzed by Sandmo (1971), Newbery and Stiglitz (1981) and many subsequent writers. In particular, the value of information in problems of this kind, and the impact of information provision on the firm's actions, has recently been analyzed by Lehmann (1988), Ormiston and Schlee (1992, 1993), Athey (forthcoming), Athey and Levin (2000) and Athey (2000). In common with most analysis of production under uncertainty, these studies focus on the case of a scalar decision variable, and approach matters analytically in terms of primal objective functions rather than in terms of indirect objective functions.

This paper examines the impact of information provision for producers facing price and production uncertainty using an Arrow-Debreu state-contingent specification of technology that admits each of the above models as a special case. The central observation of this paper is that decision making for individuals with linear-in-probabilities (that is, expected utility) preferences can always be viewed as equivalent to decisionmaking over closed, convex sets. Hence, under the presumption of an Arrow-Debreu state-contingent technology defined for a finite state space, standard tools from duality and revealed preference theory can be used to study such decisions. This permits examination of the value of information and the impact of information provision in their most natural terms, which is the space of probability distributions, much in the same manner that the dual approach allows an exhaustive examination of the theory of the price-taking firm in price space.

Viewed from this perspective, it is immediate that the value of information for individuals with linear-in-probability preferences is determined by the curvature properties of the convex hull of the individual's feasible state-contingent utilities. In particular, information is valueless in the neighborhood of kinks in the convex hull's surface and most valuable in the neighborhoods of the surface that approach linearity.

In what follows, we first introduce the model which includes specification of a state-contingent production technology and producer preferences over stochastic outcomes. We use those specifications to deduce the properties of a utility correspondence induced by the composition of the two. We then show that it can be appropriately convexified and used as the basis for the study of decisionmaking.

Then we marshal standard arguments from convex analysis and the theory of monotone comparative statics to provide a general framework for the study of the value of information to producers facing a stochastic technology. Issues addressed include the investment responses of firms to the provision of information, a general method for computing the value of information which affords easy calculation of bounds on the value of information and an easy method for determining when information is valueless, a comparison of the value of a common information structure across agents with different preferences and stochastic technologies, and a general method using Gateaux differentials for assessing the value of different information structures to an individual agent.

2 Technology and Information in a State-Contingent Framework

2.1 The state space

The state space is given by a set $\Omega = S \times N$, which allows for S possible events relevant to production and N possible signals. A signal n, typically, will be taken to to correspond to a partition of Ω according to the events $\{1_n, ..., S_n\}$. More generally, one can think of this partition into N signals as the 'finest' possible partition of the state space. *Coarser* partitions of the state space can be considered by considering the remaining elements of the power set of $\{1, 2, ..., N\}$. The entire set of partitions can be ordered by inclusion, and we will denote the corresponding ordering $\preceq_C$, where $A \preceq_C B$ is read as A is coarser than B.

2.2 Event-contingent production

Following Chambers and Quiggin (2000), the stochastic technology, which is the same for all signals,[1] is represented by an event-contingent output correspondence. Let $\mathbf{x} \in \Re^N_+$ be a vector of inputs committed prior to the resolution of uncertainty, i.e., prior to the realization of $s \in S$, and let $\mathbf{z} \in \Re^S_+$ be a vector of event-contingent outputs also chosen prior to the realization of s. Thus, if event s is realized, output z_s is produced. We confine attention to the case of a scalar output. The technology is characterized by the *event-contingent output correspondence*, which gives the vectors of event-contingent outputs that can be produced by a given vector of inputs. Formally, it is defined by:

$$Z(\mathbf{x}) = \{\mathbf{z} \in \Re^S_+ : \mathbf{x} \text{ can produce } \mathbf{z}\}.$$

The image of the output correspondence is referred to as the *output set.*

As discussed in Chambers and Quiggin (2000) we impose the following standard properties on the event-contingent output correspondence.

Properties of $Z(\mathbf{x})$ (Z)

Z.1 $\mathbf{0}_S \in Z(\mathbf{x})$ for all $\mathbf{x} \in \Re^N_+, \mathbf{z} \notin Z(\mathbf{0}_N)$ for $\mathbf{z} \neq \mathbf{0}_S$;

Z.2 $\mathbf{z}' \geq \mathbf{z} \in Z(\mathbf{x}) \Rightarrow \mathbf{z}' \in Z(\mathbf{x})$ (free disposability of outputs);

Z.3 $Z(\mathbf{x})$ is bounded for all $\mathbf{x} \in \Re^N_+$ (boundedness);

Z.4. $Z(\mathbf{x})$ is convex for all $\mathbf{x} \in \Re^N_+$;

Z.5. Z is a continuous correspondence.

[1]Hence, in this paper we do not analyze the case where the realization of the technology is random, even though actual production is.

2.3 The utility correspondence and the utility set

Now suppose that the producer is concerned to maximize an objective function of the general form $h(\mathbf{u})$ where $\mathbf{u} \in \Re^S$ and

$$u_s = u(z_s, \mathbf{x})$$

for some utility function $u : \Re_+ \times \Re_+^N \rightarrow \Re_+$, which is assumed to be continuous, concave, and monotone increasing in z_s. Define the *utility correspondence* by

$$U(\mathbf{x}) = \{\mathbf{u}(\mathbf{z}, \mathbf{x}) \in \Re_+^S : \quad \mathbf{z} \in Z(\mathbf{x})\}.$$

Lemma 1 *Assume that Z satisfies Z.1-5. Then U satisfies*

U.1 $\mathbf{u}(\mathbf{0}_S, \mathbf{x}) \in U(\mathbf{x})$ *for all* $\mathbf{x} \in \Re_+^N$;

U.2 $\mathbf{u}' \geq \mathbf{u} \in U(\mathbf{x}) \Rightarrow \mathbf{u}' \in U(\mathbf{x})$;

U.3 $U(\mathbf{x})$ *is bounded for all* $\mathbf{x} \in \Re_+^N$;

U.4 $U(\mathbf{x})$ *is convex for all* $\mathbf{x} \in \Re_+^N$;

U.5 U *is a continuous correspondence.*

The *utility set,* $\mathcal{U} \subset \Re_+^S$, is the range of the utility correspondence:

$$\mathcal{U} = \cup_{\mathbf{x}} U(\mathbf{x}).$$

Although each $U(\mathbf{x})$ is closed and convex, $\mathcal{U}$ generally is not. Denote the closure of the convex hull of $\mathcal{U}$ by

$$U^* = cl\,(conv\mathcal{U}).$$

Our discussion relies on some concepts from the theory of convex sets (Rockafellar, 1970).[2] For a convex set A, $F \subset A$ is an *exposed face* if there exists a supporting hyperplane to A, P, such that

$$P \cap A = F.$$

An exposed face that consists of a single point, i.e., $F = \{a\}$ is called an *exposed point.* $a \in A$ is an *exposed kink* if it belongs to more than one exposed face in A. For an exposed kink,

[2] We restrict attention to convex sets, $A \subseteq \Re_+^S$.

the notation $a\{B\}$ means that a belongs to the exposed faces for each of the hyperplanes in the set B.

Two convex sets A and C are *$P-$comparable* if they have intersecting exposed faces F and F', respectively, for a common supporting hyperplane P. That is, there is a P that supports A and C and that satisfies

$$P \cap A = F,$$

$$P \cap C = F',$$

and

$$F \cap F' \neq \varnothing.$$

For two $P-$comparable sets, C is a *contraction of* A if $C \subset A$.

For any convex set, A, define its *free disposal hull,* $D(A)$, by the set of points dominated by the elements of A

$$D(A) = \{b : b \leq a, \quad a \in A\}.$$

A set C is said to be a *cube* if it is the free disposal hull of a single point a, i.e.,

$$C = D(\{a\}).$$

3 Signals and probabilities

Because the state space is of the form $\Omega = S \times N$, the probability distribution may be represented by an $S \times N$ matrix P with entry p_{sn} corresponding to the joint probability of production event s and signal n. Let

$$\pi_s^0 = \sum_n p_{sn}$$

be the unconditional probability of production event s,

$$\pi_n = \sum_s p_{sn}$$

be the unconditional probability of signal realization n,[3] and

$$\pi_s^n = \frac{p_{sn}}{\pi_n}$$

[3]More properly, we should refer to this as the signal realization from the finest partition.

be the probability of event s conditional on the observation of signal realization n. Then $\boldsymbol{\pi}^0$ denotes the vector of unconditional probabilities or prior probabilities and $\boldsymbol{\pi}^n$ denotes the vector of probabilities conditional on the observation of signal realization n, or the posterior probabililties. We confine attention to the relative interior of the unit simplex, which we denote by $\mathcal{P} \subset \Re^S_{++}$. Thus, the signal is not perfectly informative regarding any event, and for all s and n, $0 < \pi^n_s < 1$. We will refer to the set of joint, prior, and posterior probabilities as an *information structure* and denote it by θ. We denote the set of prior and posterior probabilities for the information structure θ as $\Pi(\theta) = \{\boldsymbol{\pi}^0, \boldsymbol{\pi}^1, ..., \boldsymbol{\pi}^N\}$.

Given a signal with multiple possible realizations $\{1...N\}$, we can define probability vectors for *signal events*, corresponding to coarser partitions of the state space, that is, for sets of the form $\mathcal{N} \subseteq \{1...N\}$. We have

$$p^{\mathcal{N}}_s = \frac{\sum_{n\in\mathcal{N}} p_{sn}}{\sum_{n\in\mathcal{N}} \pi_n},$$

and the corresponding probability of the signal event $\mathcal{N}$ is $\sum_{n\in\mathcal{N}} \pi_n$.

4 The expected-utility function

We assume that the producer's objective function is of the expected utility or 'linear-in-probabilities' form

$$\sum_{s=1}^{S} \pi_s u(z_s, \mathbf{x}).$$

Define the *restricted expected utility function*, $V : \mathcal{P} \times \Re^N_+ \to \Re_+$, by

$$V(\boldsymbol{\pi}, \mathbf{x}) = Max \left\{ \sum_{s=1}^{S} \pi_s u_s : \mathbf{u} \in U(\mathbf{x}) \right\},$$

and the *expected utility function*, $W : \mathcal{P} \to \Re_+$, by

$$\begin{aligned} W(\boldsymbol{\pi}) &= \sup \left\{ \sum_{s=1}^{S} \pi_s u_s : \mathbf{u} \in \mathcal{U} \right\} \\ &= \sup_{\mathbf{x}} \{V(\boldsymbol{\pi}, \mathbf{x})\}. \end{aligned}$$

For convenience, we assume that solutions always exist and that $V(\boldsymbol{\pi}, \mathbf{x})$ and $W(\boldsymbol{\pi})$ are finite.

Both $V(\boldsymbol{\pi}, \mathbf{x})$ and $W(\boldsymbol{\pi})$ are convex and continuous on $\mathcal{P}$ (Färe, 1988). Denote

$$\bar{\mathbf{u}}(\boldsymbol{\pi}, \mathbf{x}) = \arg\max \left\{ \sum_{s=1}^{S} \pi_s u_s : \mathbf{u} \in U(\mathbf{x}) \right\},$$

$$\mathbf{u}(\boldsymbol{\pi}) = \arg\sup \left\{ \sum_{s=1}^{S} \pi_s u_s : \mathbf{u} \in \mathcal{U} \right\},$$

and

$$\mathbf{x}(\boldsymbol{\pi}) \in \arg\sup \{V(\boldsymbol{\pi}, \mathbf{x})\}$$

$\bar{\mathbf{u}}(\boldsymbol{\pi}, \mathbf{x})$ and $\mathbf{u}(\boldsymbol{\pi})$ are exposed faces of $U(\mathbf{x})$ and U^*, respectively. By analogy with the theory of profit maximization, in the case when they are singletons, one may think of $\bar{\mathbf{u}}(\boldsymbol{\pi}, \mathbf{x})$ and $\mathbf{u}(\boldsymbol{\pi})$ as vectors of supplies of state-contingent utilities. Moreover, we have the following analogues of Hotelling's Lemma:

$$\bar{\mathbf{u}}(\boldsymbol{\pi}, \mathbf{x}) \subseteq \partial V(\boldsymbol{\pi}, \mathbf{x})$$

and

$$\mathbf{u}(\boldsymbol{\pi}) \subseteq \partial W(\boldsymbol{\pi})$$

where $\partial f(\mathbf{x})$ denotes the subdifferential of the function $h(\mathbf{x})$ at $\mathbf{x}$.[4] Denote by $W'(\boldsymbol{\pi}; \boldsymbol{\eta})$ the one sided directional derivative (Gateaux differential) in the direction of $\boldsymbol{\eta}$:

$$W'(\boldsymbol{\pi}; \boldsymbol{\eta}) = \inf_{\varepsilon > 0} \left\{ \frac{W(\boldsymbol{\pi} + \varepsilon\boldsymbol{\eta}) - W(\boldsymbol{\pi})}{\varepsilon} \right\}.$$

$W'(\boldsymbol{\pi}; \boldsymbol{\eta})$ is positively linearly homogeneous and convex in $\boldsymbol{\eta}$, and $W'(\boldsymbol{\pi}; \boldsymbol{\eta}) \geq -W'(\boldsymbol{\pi}; -\boldsymbol{\eta})$. So long as $W(\boldsymbol{\pi})$ is finite, these Gateaux differentials always exist (Rockafellar, 1970). Moreover,

$$W'(\boldsymbol{\pi}; \boldsymbol{\eta}) = \sup_{\mathbf{u} \in \mathbf{u}(\boldsymbol{\pi})} \{\boldsymbol{\eta}\mathbf{u}\}.$$

If $W'(\boldsymbol{\pi}; \boldsymbol{\eta}) = -W'(\boldsymbol{\pi}; -\boldsymbol{\eta})$ for all $\boldsymbol{\eta}$, then $W(\boldsymbol{\pi})$ is *Gateaux differentiable* at $\boldsymbol{\pi}$, and $\mathbf{u}(\boldsymbol{\pi})$ is unique. Conversely, if $\mathbf{u}(\boldsymbol{\pi})$ is unique, then $W(\boldsymbol{\pi})$ is Gateaux differentiable at $\boldsymbol{\pi}$. Gateaux differentiability implies that $W'(\boldsymbol{\pi}; \boldsymbol{\eta})$ is linear in $\boldsymbol{\eta}$ (Rockafellar, 1970).

[4] A vector $\hat{\mathbf{x}}$ is a subgradient of a convex function, $h(\mathbf{x})$, at $\mathbf{x}$ if

$$h(\mathbf{z}) \geq h(\mathbf{x}) + \hat{\mathbf{x}}(\mathbf{z} - \mathbf{x})$$

for all $\mathbf{z}$ (Rockafellar, 1970). The subdifferential $\partial h(\mathbf{x})$ is the set of all subgradients.

Because $U(\mathbf{x})$ is convex and satisfies U.1-U.5, it may be recaptured from V by applying the duality mapping (Färe, 1988)

$$U(\mathbf{x}) = \cap_{\pi \in \mathcal{P}} \left\{ \mathbf{u} : \sum_{s=1}^{S} \pi_s u_s \leq V(\boldsymbol{\pi}, \mathbf{x}) \right\}.$$

Generally, it will not be possible to recapture $\mathcal{U}$ by a similar duality mapping. However, it does follow (Rockafellar, 1970) that

$$U^* = \cap_{\pi \in \mathcal{P}} \left\{ \mathbf{u} : \sum_{s=1}^{S} \pi_s u_s \leq W(\boldsymbol{\pi}) \right\}.$$

U^* is observationally equivalent to $\mathcal{U}$ for individuals with linear in probability preferences.

Both the restricted expected-utility function and the expected-utility function define total orderings of the probability simplex. In what follows, two particular orderings shall be useful. We say that $\hat{\boldsymbol{\pi}}$ is $(V, \mathbf{x})$ *−dominated by* $\boldsymbol{\pi}'$, denoted as $\hat{\boldsymbol{\pi}} \preceq_{(V,\mathbf{x})} \boldsymbol{\pi}'$, if $V(\boldsymbol{\pi}', \mathbf{x}) \geq V(\hat{\boldsymbol{\pi}}, \mathbf{x})$. If $\mathcal{V}$ is a class of restricted utility functions, we say that $\hat{\boldsymbol{\pi}}$ is $(\mathcal{V}, \mathbf{x})$ −dominated by $\boldsymbol{\pi}'$, denoted as $\hat{\boldsymbol{\pi}} \preceq_{(\mathcal{V},\mathbf{x})} \boldsymbol{\pi}'$, if $\hat{\boldsymbol{\pi}} \preceq_{(V,\mathbf{x})} \boldsymbol{\pi}'$ for all $V \in \mathcal{V}$. For example, consider the case where $\mathcal{V}$ consists of all V such that:

$$u_s = u(f(\mathbf{x},s)).$$

Then $\hat{\boldsymbol{\pi}} \preceq_{(\mathcal{V},\mathbf{x})} \boldsymbol{\pi}'$ is implied by requiring that the probability distribution for u induced by $\boldsymbol{\pi}'$ first-order stochastically dominates the probability distribution for u induced by $\hat{\boldsymbol{\pi}}$.

Under the supposition that u is strictly concave, then $\hat{\boldsymbol{\pi}} \preceq_{(\mathcal{V},\mathbf{x})} \boldsymbol{\pi}'$ is implied by requiring that the probability distribution for $f(\mathbf{x},s)$ induced by $\boldsymbol{\pi}'$ is less risky in the sense of Rothschild and Stiglitz than that for $\hat{\pi}$. In this case, the ordering $\preceq_{(\mathcal{V},\mathbf{x})}$ is particularly simple because the rank ordering of the utilities and the realization of the production process is independent of the producer's probabilities. More generally, the rank ordering of the event-contingent utilities will depend upon the probabilities that the producer faces. For example, if the underlying event-contingent technology has strictly convex output sets, the ordering will not apply to the entire probability simplex because the ordering of outputs and, thus, of *ex post* utilities depends directly on the probabilities faced and on the level of input utilitzation. (Chambers and Quiggin, 2000). Consequently, whether $\hat{\boldsymbol{\pi}} \preceq_{(V,\mathbf{x})} \boldsymbol{\pi}'$ or not is a result of a complex interplay of the producer's preferences and his event-contingent technology.

By analogy, we say that $\hat{\pi}$ is *W*−*dominated by* π', denoted as $\hat{\pi} \preceq_W \pi'$, if $W(\pi') \geq W(\hat{\pi})$. The following lemma links these two concepts of dominance.

Lemma 2 *$\hat{\pi} \preceq_W \pi'$ if $\hat{\pi} \preceq_{(\mathrm{V},\mathbf{x})} \pi'$ at $\mathbf{x}(\hat{\pi})$.*

Proof *Suppose $\hat{\pi} \preceq_{(\mathrm{V},\mathbf{x})} \pi'$ at $\mathbf{x}(\hat{\pi})$, then*

$$\begin{aligned} V(\pi', \mathbf{x}(\pi')) &\geq V(\pi', \mathbf{x}(\hat{\pi})) \\ &\geq V(\hat{\pi}, \mathbf{x}(\hat{\pi})). \end{aligned}$$

Corollary 3 *If $\hat{\pi} \preceq_{(\mathcal{V},\mathbf{x})} \pi'$ at $\mathbf{x}(\hat{\pi})$, then $\hat{\pi} \preceq_W \pi'$ for all $W \in \mathcal{V}$.*

Corollary 4 *If $\hat{\pi} \preceq_{(\mathrm{V},\mathbf{x})} \pi'$ for all $\mathbf{x}$, then $\hat{\pi} \preceq_W \pi'$.*

In addition to the total orderings of the simplex induced by the two expected-utility functions, we will also be interested in partial orderings of the probability simplex. An example of an ordering of the probability simplex that is often of interest is

$$\hat{\pi} \preceq_s \pi' \iff \sum_{s=1}^{k} \hat{\pi}_s \geq \sum_{s=1}^{k} \pi'_s, \quad k = 1, 2, ..., S-1.$$

If outcomes are ordered from worst to best in accordance with the indexing of states, i.e., $n > n' \Longrightarrow u_n \geq u_{n'}$, $\hat{\pi} \preceq_s \pi'$ is equivalent to stochastic dominance.

When $\hat{\pi} \preceq \pi' \Longrightarrow \hat{\pi} \preceq_{(V,\mathbf{x})} \pi'$ and $\hat{\pi} \preceq \pi' \Longrightarrow \hat{\pi} \preceq_W \pi'$,we say, respectively, that V and W *preserve* the ordering $\preceq$. When $\hat{\pi} \preceq \pi' \Longrightarrow W'(\pi'; \eta) \geq W'(\hat{\pi}; \eta)$, we say that W is *convex in the ordering* $\preceq$ *at* η. We have:

Corollary 5 *If V preserves the ordering $\preceq$ at $\mathbf{x}(\hat{\pi})$ then $\hat{\pi} \preceq \pi' \Longrightarrow \hat{\pi} \preceq_W \pi'$. If V preserves the ordering $\preceq$ for all $\mathbf{x}$ then $\hat{\pi} \preceq \pi' \Longrightarrow \hat{\pi} \preceq_W \pi'$.*

To discuss input comparative statics, we need a notion of monotonicity for sets. Consider any $A(\mathbf{t}) \subset \Re_+^N$ where $\mathbf{t}$ is drawn from a partially ordered set ordered by $\preceq$. $A(\mathbf{t})$ is said to be increasing in $\mathbf{t}$, written $\mathbf{t} \preceq \mathbf{t}' \Longrightarrow \mathbf{A}(\mathbf{t}') \geq \mathbf{A}(\mathbf{t})$, if for any $\mathbf{a}' \in \mathbf{A}(\mathbf{t}')$, $\mathbf{a} \in \mathbb{A}(\mathbf{t})$

$$\inf\{\mathbf{a}, \mathbf{a}'\} \in \mathbf{A}(\mathbf{t}) \text{ and } \sup\{\mathbf{a}, \mathbf{a}'\} \in \mathbf{A}(\mathbf{t}).$$

Given this characterization, our comparative static results on the response of input utilization to different sources of information are a direct consequence of basic results (Milgrom and Shannon, 1994; Topkis, 1998) in the theory of monotone comparative statics.

Theorem 6 *(Milgrom and Shannon) If $V(\boldsymbol{\pi}, \mathbf{x})$ is quasi-supermodular in $\mathbf{x}$ for all $\boldsymbol{\pi}$, and for all $\hat{\boldsymbol{\pi}} \preceq \boldsymbol{\pi}'$ and $\hat{\mathbf{x}} \leq \mathbf{x}'$,*

$$V(\hat{\boldsymbol{\pi}}, \hat{\mathbf{x}}) \leq (<) V(\hat{\boldsymbol{\pi}}, \mathbf{x}') \Longrightarrow V(\boldsymbol{\pi}', \hat{\mathbf{x}}) \leq (<) V(\boldsymbol{\pi}', \mathbf{x}'),$$

then $\hat{\boldsymbol{\pi}} \preceq \boldsymbol{\pi}' \Longrightarrow \mathbf{x}(\boldsymbol{\pi}') \geq \mathbf{x}(\hat{\boldsymbol{\pi}})$.

5 The value of information

We distinguish two notions of the value of information. The *value of the nth signal* and *the value of the information structure θ*. We will also distinguish between the case where the input bundle is fixed and the case where the input bundle is freely variable in response to the reception of different signals. The *value of the nth signal* for the fixed input bundle, $\mathbf{x}$, is

$$\bar{\Delta}(\boldsymbol{\pi}^n, \boldsymbol{\pi}^0, \mathbf{x}) = V(\boldsymbol{\pi}^n, \mathbf{x}) - \sup_{\mathbf{u} \in \bar{\mathbf{u}}(\boldsymbol{\pi}^0, \mathbf{x})} \left\{ \sum_s \pi_s^n u_s \right\} \geq 0,$$

where the inequality follows by the definition of V. In words, $\bar{\Delta}(\boldsymbol{\pi}^n, \boldsymbol{\pi}^0, \mathbf{x})$ is the difference between the optimal expected utility as calculated for the posterior probabilities and the producer's best alternative, using the posterior probabilities, of the production choice made for the prior probabilities.

The *value of the information structure θ for the fixed input bundle* $\mathbf{x}$ is the difference between the producer's *ex ante* expected utility when he has access to the entire information structure and his expected utility when he only knows the prior probabilities

$$\bar{I}(\theta, \mathbf{x}) = \sum_n \pi_n V(\boldsymbol{\pi}^n, \mathbf{x}) - V(\boldsymbol{\pi}^0, \mathbf{x}) \geq 0,$$

by the convexity of V because $\boldsymbol{\pi}^0 = \sum_n \pi_n \boldsymbol{\pi}^n$.[5] Notice that

$$\bar{I}(\theta, \mathbf{x}) = \sum_n \pi_n \bar{\Delta}(\boldsymbol{\pi}^n, \boldsymbol{\pi}^0, \mathbf{x}).$$

Similarly, the *value of the nth signal* is defined

$$\Delta(\boldsymbol{\pi}^n, \boldsymbol{\pi}^0) = W(\boldsymbol{\pi}^n) - \sup_{\mathbf{u} \in \mathbf{u}(\boldsymbol{\pi}^0)} \left\{ \sum_s \pi_s^n u_s \right\},$$

[5] Actually, convexity is not needed to establish that the value of information is positive. It follows immediately from the fact that in the presence of a signal, the producer could always choose $\mathbf{u}(\boldsymbol{\pi}^0, \mathbf{x})$.

and the *value of the information structure* θ is

$$\begin{aligned} I(\theta) &= \sum_n \pi_n W(\boldsymbol{\pi}^n) - W(\boldsymbol{\pi}^0) \\ &= \sum_n \pi_n \triangle(\boldsymbol{\pi}^n, \boldsymbol{\pi}^0) \geq 0. \end{aligned}$$

Our next result bounds the value of signals and the value of the information structure for a given technology and preferences.

Theorem 7

$$V(\boldsymbol{\pi}^0, \mathbf{x}) \left(\max\left\{ \frac{\pi_1^n}{\pi_1^0}, ..., \frac{\pi_S^n}{\pi_S^0} \right\} - 1 \right) \geq \bar{\triangle}(\boldsymbol{\pi}^n, \boldsymbol{\pi}^0, \mathbf{x}) \geq 0$$

$$V(\boldsymbol{\pi}^0, \mathbf{x}) \left(\sum_n \max\left\{ \frac{p_{1n}}{\pi_1^0}, ..., \frac{p_{Sn}}{\pi_S^0} \right\} - 1 \right) \geq \bar{I}(\theta, \mathbf{x}) \geq 0,$$

$$W(\boldsymbol{\pi}^0) \left(\max\left\{ \frac{\pi_1^n}{\pi_1^0}, ..., \frac{\pi_S^n}{\pi_S^0} \right\} - 1 \right) \geq \triangle(\boldsymbol{\pi}^n, \boldsymbol{\pi}^0) \geq 0,$$

$$W(\boldsymbol{\pi}^0) \left(\sum_n \max\left\{ \frac{p_{n1}}{\pi_1^0}, ..., \frac{p_{nS}}{\pi_S^0} \right\} - 1 \right) \geq I(\theta) \geq 0$$

Proof *The proof is for* $\bar{\triangle}(\boldsymbol{\pi}^n, \boldsymbol{\pi}^0, \mathbf{x})$. *An exactly parallel proof applies to the remaining expressions. By the definition of* V,

$$U(\mathbf{x}) \subset \left\{ \mathbf{u} : \sum_s \pi_s^0 u_s \leq V(\boldsymbol{\pi}^0, \mathbf{x}) \right\}.$$

Thus, $V(\boldsymbol{\pi}, \mathbf{x})$ *must be bounded from above by*

$$\begin{aligned} \bar{V}(\boldsymbol{\pi}, \mathbf{x}) &= \max\left\{ \sum_s \pi_s u_s : \sum_s \pi_s^0 u_s \leq V(\boldsymbol{\pi}^0, \mathbf{x}) \right\} \\ &= V(\boldsymbol{\pi}^0, \mathbf{x}) \max\left\{ \frac{\pi_1}{\pi_1^0}, ..., \frac{\pi_S}{\pi_S^0} \right\}. \end{aligned}$$

From the proof of Theorem 7, one observes that signals and information generally are the most valuable when the expected utility function takes the form $W(\boldsymbol{\pi}^0) \max\left\{ \frac{\pi_1}{\pi_1^0}, ..., \frac{\pi_S}{\pi_S^0} \right\}$. This corresponds to the case where U^* has a single exposed face. In this case, where the frontier of U^* is linear, even the smallest changes in probability distributions will lead to large changes in choice of state-contingent utilities and hence in state-contingent outputs.

In other words, the producer engages in plunging behavior. On the other hand, signals and information are least valuable when the expected utility function is linear over a range of probability distributions. By standard results from duality theory (McFadden, 1978), linearity in the expected utility function maps directly into kinks in the frontier of U^*. We pursue this point formally in the following series of theorems and corollaries.

Theorem 8 $\bar{\triangle}(\boldsymbol{\pi}^n, \boldsymbol{\pi}^0, \mathbf{x}) = 0$ *if and only if* $\bar{\mathbf{u}}(\boldsymbol{\pi}^0, \mathbf{x})$ *contains an exposed kink in* $U(\mathbf{x})$ *for* $(\boldsymbol{\pi}^n, \boldsymbol{\pi}^0)$. $\bar{I}(\theta, \mathbf{x}) = 0$ *if and only if* $U(\mathbf{x})$ *is a contraction of*

$$\cap_{\pi \in \mathbf{\Pi}(\theta)} \left\{ \mathbf{u} : \sum_s \pi_s u_s \leq \sup \left\{ \sum_s \pi_s \bar{u}_s \left(\boldsymbol{\pi}^0, \mathbf{x} \right) \right\} \right\}.$$

Proof: Suppose

$$V(\boldsymbol{\pi}^n, \mathbf{x}) - \sup \left\{ \sum_s \pi_s^n \bar{u}_s \left(\boldsymbol{\pi}^0, \mathbf{x} \right) \right\} = 0.$$

Hence, $\bar{\mathbf{u}}(\boldsymbol{\pi}^0, \mathbf{x})$ *must contain at least one element that belongs to faces of* $U(\mathbf{x})$ *for both* $\boldsymbol{\pi}^0$ *and* $\boldsymbol{\pi}^n$. *Conversely, if* $\bar{\mathbf{u}}(\boldsymbol{\pi}^0, \mathbf{x})$ *contains an exposed kink in* $U(\mathbf{x})$ *for* $(\boldsymbol{\pi}^n, \boldsymbol{\pi}^0)$, $\bar{\triangle}(\boldsymbol{\pi}^n, \boldsymbol{\pi}^0, \mathbf{x})$ *cannot be strictly positive. For the second part, suppose to the contrary that there exists a feasible* $\mathbf{u}$ *such that*

$$\sum_s \pi_s^n u_s > \sup \left\{ \sum_s \pi_s^n \bar{u}_s \left(\boldsymbol{\pi}^0, \mathbf{x} \right) \right\}$$

for any n. *Then* $\bar{I}(\theta, \mathbf{x}) > 0$.

Corollary 9 *Suppose* $\bar{\mathbf{u}}(\boldsymbol{\pi}^0, \mathbf{x})$ *is unique, then* $\bar{I}(\theta, \mathbf{x}) = 0$ *if and only if* $\bar{\mathbf{u}}(\boldsymbol{\pi}^0, \mathbf{x})$ *is an exposed kink for* $\Pi(\theta)$.

Corollary 10 $\bar{I}(\theta, \mathbf{x}) = 0$ *for all information structures* θ *if and only if* $U(\mathbf{x})$ *is a cube.*

Exactly parallel arguments reveal:

Theorem 11 $\triangle(\pi^n, \pi^0) = 0$ *if and only if* $\mathbf{u}(\pi^0)$ *contains an exposed kink in* U^* *for* (π^n, π^0). $I(\theta) = 0$ *if and only if* U^* *is a contraction of*

$$\cap_{\pi \in \mathbf{\Pi}(\theta)} \left\{ \mathbf{u} : \sum_s \pi_s u_s \leq \sup \left\{ \sum_s \pi_s u_s \left(\pi^0 \right) \right\} \right\}.$$

Corollary 12 Suppose W is Gateaux differentiable at $\boldsymbol{\pi}^0$, then $I(\theta) = 0$ if and only if $\mathbf{u}(\boldsymbol{\pi}^0)$ is an exposed kink for $\Pi(\theta)$.

Corollary 13 $I(\theta) = 0$ for all possible information structures if and only if U^* is a cube.

When W is not Gateaux differentiable at $\boldsymbol{\pi}^0$, there is not an unique optimal solution for the prior probabilities. Hence, in that case the exposed face of U^* must be at least locally linear. If either a signal or an information structure is to be valueless, however, the hyperplanes defined by the associated posterior distributions must pass through at least one exposed kink in the exposed face. When this happens, varying the posterior probabilities brings no adjustment on the part of the producer. However, when W is Gateaux differentiable at $\boldsymbol{\pi}^0$, information can only be valueless if the optimal solution for the prior distribution corresponds to an exposed kink for the entire range of posterior probabilities.

We conclude this discussion on the general value of information with

Corollary 14 *For an information structure θ, the following are equivalent*

(i) $I(\theta) = 0$;

(ii) W is linear on $\Pi(\theta)$;

(iii) For all $\boldsymbol{\pi}^n \in \Pi(\theta)$, $\Delta(\boldsymbol{\pi}^n, \boldsymbol{\pi}^0) = 0$.

Conversely, the upper bound in Theorem 7 is attained when

$$U^* = \left\{ \mathbf{u} : \sum_s \pi_s^0 u_s \leq W\left(\boldsymbol{\pi}^0\right) \right\}.$$

5.1 Comparing Different Valuations of the Same Information Structure

We now turn our attention to making comparisons of the value of the same information structure across two different utility correspondences, U^1, U^2 with associated U^{1*}, U^{2*}. To conserve space, we only state results in terms of U^{1*}, U^{2*}. Exactly parallel results apply for the utility correspondences, and we leave the restatement of those results to the reader. We begin with an invariance result.

Lemma 15 *Suppose*

$$U^{2*} = \hat{\mathbf{u}} + \lambda U^{1*}$$

for fixed $\hat{\mathbf{u}} \in \Re^S_+$ *and* $\lambda \geq 0$, *then*

$$\triangle^2 \left(\boldsymbol{\pi}^n, \boldsymbol{\pi}^0\right) = \lambda \triangle^1 \left(\boldsymbol{\pi}^n, \boldsymbol{\pi}^0\right).$$

By this lemma, it follows immediately that we can assume, without true loss of generality, that $\mathbf{u}^1 (\boldsymbol{\pi}^0) \cap \mathbf{u}^2 (\boldsymbol{\pi}^0) \neq \varnothing$, since, given $\mathbf{u} \in \mathbf{u}^1 (\boldsymbol{\pi}^0), \tilde{\mathbf{u}} \in \mathbf{u}^2 (\boldsymbol{\pi}^0)$, we can 'attach' the two sets by adding $\hat{\mathbf{u}} = \mathbf{u} - \tilde{\mathbf{u}}$ to the original U^{2*}. Consequently,

Theorem 16 *Suppose that for* U^{1*}, U^{2*}, $\mathbf{u}^1 (\boldsymbol{\pi}^0) \cap \mathbf{u}^2 (\boldsymbol{\pi}^0) \neq \varnothing$, *then*

$$\triangle^1 \left(\boldsymbol{\pi}^n, \boldsymbol{\pi}^0\right) \geq \triangle^2 \left(\boldsymbol{\pi}^n, \boldsymbol{\pi}^0\right)$$

if and only U^{2*} *is a contraction of*

$$\left\{\mathbf{u} : \sum_s \pi^n_s u_s \leq W^1 \left(\boldsymbol{\pi}^n\right)\right\} \cap \left\{\mathbf{u} : \sum_s \pi^0_s u_s \leq W^1 \left(\boldsymbol{\pi}^0\right)\right\}.$$

Proof *By hypothesis* $W^1 (\boldsymbol{\pi}^0) = W^2 (\boldsymbol{\pi}^0)$ *and* $W^1 (\boldsymbol{\pi}^n) \geq W^2 (\boldsymbol{\pi}^n)$.

Corollary 17 *Suppose that for* U^{1*}, U^{2*}, $\mathbf{u}^1 (\boldsymbol{\pi}^0) \cap \mathbf{u}^2 (\boldsymbol{\pi}^0) \neq \varnothing$, *then* $\triangle^1 (\boldsymbol{\pi}^n, \boldsymbol{\pi}^0) \geq \triangle^2 (\boldsymbol{\pi}^n, \boldsymbol{\pi}^0)$ *for all* n *if and only if* U^{2*} *is a contraction of* $\cap_{\boldsymbol{\pi} \in \boldsymbol{\Pi}(\theta)} \{\mathbf{u} : \sum_s \pi_s u_s \leq W^1 (\boldsymbol{\pi})\}$.

Corollary 18 *Suppose that for* U^{1*}, U^{2*}, $\mathbf{u}^1 (\boldsymbol{\pi}^0) \cap \mathbf{u}^2 (\boldsymbol{\pi}^0) \neq \varnothing$, *then*

$$I^1 (\theta) \geq I^2 (\theta)$$

for all information structures θ *sharing the common prior probability distribution* $\boldsymbol{\pi}^0$ *if and only if* U^{2*} *is a contraction of* U^{1*}.

Corollary 19 *Suppose that for* U^{1*}, U^{2*}, $\mathbf{u}^1 (\boldsymbol{\pi}^0) \cap \mathbf{u}^2 (\boldsymbol{\pi}^0) \neq \varnothing$ *with* $I^1 (\theta) \geq I^2 (\theta)$ *for all information structures* θ *sharing the common prior probability distribution* $\boldsymbol{\pi}^0$, *then*

$$\sum_n \pi_n \left(W^1 \left(\boldsymbol{\pi}^n\right) - W^D \left(\boldsymbol{\pi}^n\right)\right) \geq I^1 (\theta) - I^2 (\theta),$$

where $W^D (\boldsymbol{\pi}) = \sup \{\sum_s \pi_s u_s : \mathbf{u} \in D (\mathbf{u}^1 (\boldsymbol{\pi}^0) \cap \mathbf{u}^2 (\boldsymbol{\pi}^0))\}$.

5.1.1 Global comparisons and the marginal value of information

By lemma 15, the assumption that

$$\mathbf{u}^1\left(\pi^0\right) \cap \mathbf{u}^2\left(\pi^0\right) \neq \varnothing$$

involves no loss of generality, since given $\mathbf{u}^0 \in \mathbf{u}^1\left(\pi^0\right), \tilde{\mathbf{u}}^0 \in \mathbf{u}^2\left(\pi^0\right)$, we can 'attach' the two sets by adding $\hat{\mathbf{u}} = \mathbf{u} - \tilde{\mathbf{u}}$ to the original U^{2*}. This attachment can be done for any $\boldsymbol{\pi}$. This motivates: U^{2*} is a *global contraction* of U^{1*} if for any π, there exist $\mathbf{u} \in \mathbf{u}^1(\boldsymbol{\pi}), \tilde{\mathbf{u}} \in \mathbf{u}^2(\boldsymbol{\pi})$ such that $U^{2*} + (\mathbf{u} - \tilde{\mathbf{u}})$ is a contraction of U^{1*}. We have

Corollary 20 *$I^1(\theta) \geq I^2(\theta)$ for all information structures θ, regardless of the prior $\boldsymbol{\pi}$, if and only if U^{2*} is a global contraction of U^{1*}.*

More significantly, the global contraction relationship allows us to order utility correspondences with the respect to the marginal value of information. Suppose $\theta \preceq_C \theta'$. Then, as discussed above, the information structure θ' can be considered as the observation of one or more signals in addition to that associated with the information structure θ. If U^{2*} is a global contraction of U^{1*} then, conditional on any particular signal realization π^n for θ, the value of the additional signal is greater under U^{1*} than under U^{2*}, and hence the marginal value of information is greater for U^{1*} than for U^{2*}.

More formally, we have

Theorem 21 *$I^1(\theta') - I^1(\theta^o) \geq I^2(\theta') - I^2(\theta^o)$ whenever $\theta^o \preceq_C \theta'$ if and only if U^{2*} is a global contraction of U^{1*}*

5.2 Evaluating Different Information Structures

In preceding sections, we have examined the impact that the technology and preferences play in determining an individual producer's valuation of information. Because decisionmakers routinely are faced with alternative sources of information (for example, competing forecasting services), it is natural to ask what makes one information source more valuable to the risk-averse firm than another. The answer lies in both the structure of technology and the structure of preferences as manifested by the utility correspondence.

Since the work of Blackwell, this problem has been well understood when two information structures can be ordered by $\preceq_C$. In our set-up, the basic result due to Blackwell (1951) is a trivial consequence of the convexity properties of W. Consider, for example, the information structure given by θ and then compare it with the next coarsest partition of the state space, i.e., one where two signals are compressed into a signal event. Let the signals that constitute the signal event be defined by $\mathcal{N} = \{k, k'\}$. Then the difference between the value of information for θ and this coarser partition of the state space is

$$\pi_k W\left(\boldsymbol{\pi}^k\right) + \pi_{k'} W\left(\boldsymbol{\pi}^{k'}\right) - \left(\pi_k + \pi_{k'}\right) W\left(\mathrm{p}^{\mathcal{N}}\right)$$

which has the same sign as

$$\frac{\pi_k}{\pi_k + \pi_{k'}} W\left(\boldsymbol{\pi}^k\right) + \frac{\pi_{k'}}{\pi_k + \pi_{k'}} W\left(\boldsymbol{\pi}^{k'}\right) - W\left(\mathrm{p}^{\mathcal{N}}\right) \geq 0,$$

where the inequality follows by convexity of W and the fact that

$$\mathrm{p}^{\mathcal{N}} = \frac{\pi_k}{\pi_k + \pi_{k'}} \boldsymbol{\pi}^k + \frac{\pi_{k'}}{\pi_k + \pi_{k'}} \boldsymbol{\pi}^{k'}.$$

By induction, therefore, finer partitions of the state space are always preferred to coarser ones. In what follows, therefore, we restrict attention to structures that cannot be ranked by $\preceq_C$.

In general, $I\left(\theta^1\right) \geq I\left(\theta^2\right)$ if and only if

$$\sum_n \pi_n^1 \triangle\left(\boldsymbol{\pi}^{1n}, \boldsymbol{\pi}^{10}\right) \geq \sum_n \pi_n^2 \triangle\left(\boldsymbol{\pi}^{2n}, \boldsymbol{\pi}^{20}\right),$$

where $\boldsymbol{\pi}^{in}$ denotes the probability distribution conditional on the observation of signal realization n under information structure i, and π_n^i denotes the probability of the signal realization n under information structure i. It is immediate that a sufficient condition for $I\left(\theta^1\right) \geq I\left(\theta^2\right)$ is that the random variable $\triangle$ is stochastically larger under θ^1 than under θ^2. Define

$$F^{\triangle}(\delta, \theta) = \left\{\sum_k \pi_k : \triangle\left(\boldsymbol{\pi}^k, \boldsymbol{\pi}^0\right) \leq \delta, \quad \left(\pi_k, \boldsymbol{\pi}^k, \boldsymbol{\pi}^0\right) \in \theta\right\},$$

with corresponding density

$$f^{\triangle}(\delta, \theta) = \left\{\sum_k \pi_k : \triangle\left(\boldsymbol{\pi}^k, \boldsymbol{\pi}^0\right) = \delta, \quad \left(\pi_k, \boldsymbol{\pi}^k, \boldsymbol{\pi}^0\right) \in \theta\right\}.$$

It then follows immediately that $F^{\triangle}(\delta,\theta^2) \geq F^{\triangle}(\delta,\theta^1)$ for all δ implies that $I(\theta^1) \geq I(\theta^2)$. By basic results on stochastic majorization (Marshall and Olkin, 1979),[6] it is also true that

$$\frac{f^{\triangle}(\delta,\theta^2)}{1-F^{\triangle}(\delta,\theta^2)} \geq \frac{f^{\triangle}(\delta,\theta^1)}{1-F^{\triangle}(\delta,\theta^1)}$$

for all $\delta \geq 0$ implies $I(\theta^1) \geq I(\theta^2)$, where $\frac{f^{\triangle}(\delta,\theta)}{1-F^{\triangle}(\delta,\theta)}$ is the hazard rate. The hazard rate condition has the standard interpretation that bad news, regarding the realizations of $\triangle$, is more likely under θ^2 than under θ^1.

In what follows, we will focus on the the case of a common prior where the probability distribution for the signal realizations is also common ($\pi_n^1 = \pi_n^2$ for all n) to both information structures. In that case, $I(\theta^1) \geq I(\theta^2)$, if and only if

$$\sum_n \pi_n \left(W\left(\boldsymbol{\pi}^{1n}\right) - W\left(\boldsymbol{\pi}^{2n}\right)\right) \geq 0. \tag{1}$$

There are at least two approaches that can be taken here. The first, which parallels the literature on majorization and convexity, is to determine conditions under which (1) holds for all W convex. The second approach is restrict attention to narrower classes of W and then to determine conditions under which (1) holds for all members of that class.

We first pursue the former and deduce some straightforward consequences of the convexity properties of W. These results apply to the entire class of convex expected utility functions, and thus to the entire class of linear-in-probabilities specifications. Consequently, they can be recognized as specific manifestations of the classic results of Blackwell.

Theorem 22 *Suppose θ^1 and θ^2 share a common prior and probability distributions for the signal realizations. If* $\boldsymbol{\pi}^{2n} = \lambda\boldsymbol{\pi}^0 + (1-\lambda)\boldsymbol{\pi}^{1n}, \quad n = 1,2,...,N, \quad 0 < \lambda < 1,$

$$I(\theta^2) \leq (1-\lambda) I(\theta^1).$$

Each member of θ^2 lies between the corresponding posterior in θ^1 and the prior distribution. Hence, θ^1, by analogy with the literature on risk orderings one might say that, θ^2 is[7] a *multiplicative compression of* θ^1. Because the expected utility function is convex in probabilities, there is always a loss associated with compressing the distribution of posteriors.

[6]Here the restriction that $u \in \Re_+^S$ is crucial.

[7]Even more accurately, it is a multiplicative compression towards the common prior.

Multiplicative compressions require all posteriors from the compressed distribution to bear the same relation to the prior and the posterior from the original distribution. However, by analogy with the literatures on majorization and risk orderings, it is now apparent how one can start from a particular information structure and proceed through a finite series of steps to construct another information structure that dominates the original. In particular, suppose we consider two information structures that are identical apart from the posteriors for two signal realizations. For convenience let those two signal realizations be indexed by 1 and 2, respectively. Now suppose that

$$\boldsymbol{\pi}^{21} = \lambda \boldsymbol{\pi}^{11} + (1 - \lambda)\, \boldsymbol{\pi}^{0},$$

and

$$\boldsymbol{\pi}^{22} = \lambda \boldsymbol{\pi}^{12} + (1 - \lambda)\, \boldsymbol{\pi}^{0}, \quad 0 < \lambda < 1.$$

By convexity of the expected utility function, it follows immediately that

$$\lambda \left(W\left(\boldsymbol{\pi}^{11}\right) - W\left(\boldsymbol{\pi}^{0}\right)\right) \geq W\left(\boldsymbol{\pi}^{22}\right) - W\left(\boldsymbol{\pi}^{0}\right),$$

and

$$\lambda \left(W\left(\boldsymbol{\pi}^{12}\right) - W\left(\boldsymbol{\pi}^{0}\right)\right) \geq W\left(\boldsymbol{\pi}^{21}\right) - W\left(\boldsymbol{\pi}^{0}\right),$$

whence

$$I\left(\theta^{1}\right) \geq I\left(\theta^{2}\right).$$

Again borrowing terminology from the literature on risk orderings and majorization, we could say that $(\boldsymbol{\pi}^{21}, \boldsymbol{\pi}^{22})$ is a *pairwise compression* of $(\boldsymbol{\pi}^{11}, \boldsymbol{\pi}^{12})$. Because we can repeat this process to induce another information structure, θ^3, which is a pairwise compression away from θ^2, it follows by induction that

Theorem 23 *If θ^k can be arrived at by a chain of pairwise compressions from θ^1, then $I\left(\theta^1\right) \geq I\left(\theta^k\right)$.*

Theorems 22 and 23 demonstrate that one can always define general orderings of the simplex, $\preceq$, which always correspond to an increase in the value of information. We now pursue a general procedure for generating such orderings using Gateaux differentials. To proceed notice that the assumption that θ^1 and θ^2 have common priors implies that one can think of the differences between the two sets of posterior distributions being expressible as

$$\boldsymbol{\pi}^{1n} - \boldsymbol{\pi}^{2n} = \boldsymbol{\varepsilon}^n, n = 1, 2, ..., N$$

subject to the restrictions $\sum_n \pi_n \boldsymbol{\varepsilon}^n = \mathbf{0}_S$ and $\sum_s \varepsilon_s^n = 0$. A direct consequence of $\mathbf{u}(\boldsymbol{\pi}) \subseteq \partial W(\boldsymbol{\pi})$ applied to (1) is

Lemma 24 *Suppose θ^1 and θ^2 share probability distributions for the signal realizations. If*

$$\boldsymbol{\pi}^{1n} - \boldsymbol{\pi}^{2n} = \boldsymbol{\varepsilon}^n, n = 1, 2, ..., N$$

with $\sum_n \pi_n \boldsymbol{\varepsilon}^n = \mathbf{0}_S$ and $\sum_s \varepsilon_s^n = 0$ then

$$\sum_n \pi_n \inf_{\mathbf{u} \in \mathbf{u}(\boldsymbol{\pi}^{1n})} \{\boldsymbol{\varepsilon}^n \mathbf{u}\} \geq I(\theta^1) - I(\theta^2) \geq \sum_n \pi_n \sup_{\mathbf{u} \in \mathbf{u}(\boldsymbol{\pi}^{2n})} \{\boldsymbol{\varepsilon}^n \mathbf{u}\}.$$

The inequality in the lemma can rewritten in terms of directional derivatives as

$$-\sum_n \pi_n W'(\boldsymbol{\pi}^{1n}; -\boldsymbol{\varepsilon}^n) \geq I(\theta^1) - I(\theta^2) \geq \sum_n \pi_n W'(\boldsymbol{\pi}^{2n}; \boldsymbol{\varepsilon}^n)$$

Thus, a necessary condition for information structure θ^1 to dominate θ^2 is that given the productive response to θ^1 its replacement by θ^2 cannot, on average, constitute 'good news', while a sufficient condition is that, given the productive response to information structure θ^2, its replacemement by θ^1 constitutes, on average,'good news'. The reasoning behind the sufficient condition is transparent. If replacing θ^2 by θ^1 is, on average, good news given the productive response to θ^2, then it must be even better news when the producer adjusts optimally to θ^1.

In general, any two information structures with common prior and probability distributions may be linked by a sequence of information structures having all the posterior distributions in common except for two. Lemma 24 is perhaps most transparent in this case. Let those two, for the sake of convenience be $n = 1, 2$. Then Lemma 24, implies that $I(\theta^1) \geq I(\theta^2)$ if

$$W'(\boldsymbol{\pi}^{22}; \boldsymbol{\varepsilon}^2) \geq -W'(\boldsymbol{\pi}^{21}; -\boldsymbol{\varepsilon}^2).$$

When $\boldsymbol{\varepsilon}^2$ is suitably small, this condition is necessary and sufficient. If $\boldsymbol{\pi}^{21} \preceq \boldsymbol{\pi}^{22}$, then convexity of W in the order $\preceq$ in the direction of $\boldsymbol{\varepsilon}^2$ guarantees $I(\theta^1) \geq I(\theta^2)$.

More generally, however, we conclude

Theorem 25 *Consider the partial ordering of the probability simplex* $\boldsymbol{\pi}^{22} \preceq_{(\pi^{21})} \boldsymbol{\pi}^{12}$ *defined by*

$$\boldsymbol{\pi}^{12} \in \left\{ \boldsymbol{\pi} : W' \left(\boldsymbol{\pi}^{22}; \boldsymbol{\pi} - \boldsymbol{\pi}^{22} \right) + W' \left(\boldsymbol{\pi}^{21}; \boldsymbol{\pi}^{22} - \boldsymbol{\pi} \right) \geq 0 \right\}.$$

For a (small) mean preserving change of the posterior distribution of the form

$$\varepsilon^1 = \left(\boldsymbol{\pi}^{11} - \boldsymbol{\pi}^{21} \right) = \frac{\pi_2}{\pi_1} \left(\boldsymbol{\pi}^{12} - \boldsymbol{\pi}^{22} \right)$$

$I\left(\theta^1\right) \geq I\left(\theta^2\right)$ *if and only if* $\boldsymbol{\pi}^{22} \preceq_{(\pi^{21})} \boldsymbol{\pi}^{12}$.

Theorem 25 offers a general method for generating more informative information structures. It follows trivially that any two information structures which can be related via these type of mean preserving changes can also be ranked directly in terms of informativeness.

We now turn our attention to alternative orderings of the probability simplex, and the role that they can play in ranking the informativeness of information structures. Suppose that

$$\boldsymbol{\pi}^n \preceq \boldsymbol{\pi}^{n'},$$

for an arbitrary ordering of the simplex. Then choose $\boldsymbol{\delta}$ so that $\sum_s \delta_s = 0$. We will refer to $\left(\boldsymbol{\pi}^n - \boldsymbol{\delta}, \boldsymbol{\pi}^{n'} + \frac{\pi_n}{\pi_{n'}} \boldsymbol{\delta} \right)$ as a *mean preserving spread of the ordered pair* $\left(\boldsymbol{\pi}^n, \boldsymbol{\pi}^{n'} \right)$ if

$$\boldsymbol{\pi}^n - \boldsymbol{\delta} \preceq \boldsymbol{\pi}^n \preceq \boldsymbol{\pi}^{n'} \preceq \boldsymbol{\pi}^{n'} + \frac{\pi_n}{\pi_{n'}} \boldsymbol{\delta}.$$

Lemma 26 *If* $W(\boldsymbol{\pi})$ *is convex in the ordering* $\preceq$ *at* $\boldsymbol{\delta}$*, the ex ante value of the posterior distribution increases as a result of a (small) mean preserving spread of the ordered pair*$\left(\boldsymbol{\pi}^n \preceq \boldsymbol{\pi}^{n'} \right)$.

Proof Create the mean preserving spread of the ordered pair $\boldsymbol{\pi}^n \preceq \boldsymbol{\pi}^{n'}$ given by $\alpha > 0$ but arbitrarily small:

$$\left(\boldsymbol{\pi}^n - \alpha \boldsymbol{\delta}, \boldsymbol{\pi}^{n'} + \alpha \frac{\pi_n}{\pi_{n'}} \boldsymbol{\delta} \right).$$

The variation in the value of information is

$$\pi_n W' \left(\boldsymbol{\pi}^n; -\boldsymbol{\delta} \right) + \pi_{n'} W' \left(\boldsymbol{\pi}^{n'}; \frac{\pi_n}{\pi_{n'}} \boldsymbol{\delta} \right),$$

which by the positive linear homogeneity of W' reduces to

$$\pi_n \left(W' \left(\boldsymbol{\pi}^n; -\boldsymbol{\delta} \right) + W' \left(\boldsymbol{\pi}^{n'}; \boldsymbol{\delta} \right) \right).$$

Convexity in the order at $\boldsymbol{\delta}$ and the basic properties of W' imply

$$W'\left(\boldsymbol{\pi}^{n'};\boldsymbol{\delta}\right) \geq W'\left(\boldsymbol{\pi}^{n};\boldsymbol{\delta}\right) \geq -W'\left(\boldsymbol{\pi}^{n};-\boldsymbol{\delta}\right).$$

It follows immediately from the proof of Lemma 26 that

Corollary 27 *If W is Gateaux differentiable at $\boldsymbol{\pi}^n$, the ex ante value of the posterior distribution increases as a result of a (small) mean preserving spread of the ordered pair$(\boldsymbol{\pi}^n \preceq \boldsymbol{\pi}^{n'})$ if and only if W is convex in the ordering $\preceq$ at $\boldsymbol{\delta}$.*

By analogy with the arguments made concerning multiplicative compressions, we now have

Theorem 28 *Suppose that θ^1 can be arrived at by a chain of (small) mean preserving spreads from θ^2, then if W is convex in the order $\preceq$ at the mean preserving spreads, $I\left(\theta^1\right) \geq I\left(\theta^2\right)$.*

Lemma 26 starts from the presumption that $\boldsymbol{\pi}^n \preceq \boldsymbol{\pi}^{n'}$, chooses a direction of movement that is associated with mean preserving spread around that pair, and then shows that convexity in the ordering is sufficient for the producer to prefer the more dispersed distribution (in terms of the ordering) to the original distribution. An alternative approach, largely due to Athey and Levin (2000), is to use an ordered pair to define the direction of movement, and then to identify posterior distributions from which movement in the direction of the ordered pair enhances the informativeness of the information structure.

Suppose we have a general ordering $\preceq$ on the probability simplex, and that we pick two distributions ordered so that $\boldsymbol{\pi}^o \preceq \boldsymbol{\pi}^*$. Now consider the mean preserving change of two posteriors $\boldsymbol{\pi}^n$ and $\boldsymbol{\pi}^{n'}$ in the direction of $(\boldsymbol{\pi}^* - \boldsymbol{\pi}^o)$, i.e.,

$$\boldsymbol{\varepsilon}^{n'} = \alpha\left(\boldsymbol{\pi}^* - \boldsymbol{\pi}^o\right),$$

$$\boldsymbol{\varepsilon}^{n} = \frac{\pi_{n'}}{\pi_n}\alpha\left(\boldsymbol{\pi}^o - \boldsymbol{\pi}^*\right).$$

for $\alpha > 0$, but arbitrarily small. In the limit, the associated variation in $I(\theta)$ is proportional to:

$$W'\left(\boldsymbol{\pi}^{n'};\boldsymbol{\pi}^* - \boldsymbol{\pi}^o\right) + W'\left(\boldsymbol{\pi}^{n};\boldsymbol{\pi}^o - \boldsymbol{\pi}^*\right).$$

By analogy with Theorem 25 and Lemma 26, it is, therefore, immediate that

Theorem 29 *Define the partial ordering of the probability simplex* $\boldsymbol{\pi}^n \preceq_{\preceq} \boldsymbol{\pi}^{n'}$ *by*

$$\left(\boldsymbol{\pi}^{n'}, \boldsymbol{\pi}^{n}\right) \in \cap_{\boldsymbol{\pi}^o \preceq \boldsymbol{\pi}^*} \left\{(\hat{\boldsymbol{\pi}}, \tilde{\boldsymbol{\pi}}) : W'\left(\hat{\boldsymbol{\pi}}; \boldsymbol{\pi}^* - \boldsymbol{\pi}^o\right) + W'\left(\tilde{\boldsymbol{\pi}}; \boldsymbol{\pi}^o - \boldsymbol{\pi}^*\right) \geq 0\right\}$$

Then a small mean preserving change

$$\boldsymbol{\varepsilon}^{n'} = \alpha\left(\boldsymbol{\pi}^* - \boldsymbol{\pi}^o\right),$$

$$\boldsymbol{\varepsilon}^{n} = \frac{\pi_{n'}}{\pi_n}\alpha\left(\boldsymbol{\pi}^o - \boldsymbol{\pi}^*\right).$$

with $\boldsymbol{\pi}^o \preceq \boldsymbol{\pi}^*$ *increases the informativeness of* $I(\theta)$ *if*

$$\boldsymbol{\pi}^n \preceq_{\preceq} \boldsymbol{\pi}^{n'}.$$

An immediate consequence of Theorem 29 is that any two information sets which can be related by a series of such mean preserving changes can be ranked according to informativeness. However, following Athey and Levin (2000), one can deduce even more. Suppose that for both θ^1 and θ^2, the signal realizations are ordered so that

$$\boldsymbol{\pi}^{i1} \preceq_{\preceq} \boldsymbol{\pi}^{i2} ... \preceq_{\preceq} \boldsymbol{\pi}^{iN}.$$

The information structures can then be ranked if the following Monotone Information Ordering (MIO), defined by Athey and Levin (2000) is satisfied

$$\mathbf{p}^{1\mathfrak{n}} \preceq_{\preceq} \mathbf{p}^{2\mathfrak{n}}, n = 1...N-1 \tag{2}$$

where $\mathbf{p}^{i\mathfrak{n}}$ is the probability vector associated with the signal event $n = \{1...n\}$ corresponding to the observation of one of the worst n signals, so that

$$p_s^{i\mathfrak{N}} = \frac{\sum_{i=1}^{n} p_{sn}}{\sum_{i=1}^{n} \pi_n}$$

That is, the signal realizations of the more informative information structure are 'more spread out' with respect to $\preceq$, exactly as in the Monotone Information Ordering of Athey and Levin (2000).

Theorem 30 *Suppose* θ^1 *and* θ^2 *share a common prior and probability distributions for the signal realizations. Then, if* θ^1 *and* θ^2 *satisfy the Monotone Information Ordering condition (2) for* $\preceq_{\preceq}$, $I(\theta^1) \geq I(\theta^2)$.

To relate the general approaches in Theorems 28 and 30, it is perhaps most instructive to restrict attention to the case (implicitly considered by most authors) where W is everywhere Gateaux differentiable. Convexity in the order $\preceq$ at $\boldsymbol{\delta}$ then requires for $\boldsymbol{\pi}^n \preceq \boldsymbol{\pi}^{n'}$

$$W'\left(\pi^{n'};\delta\right) - W'\left(\pi^n;\delta\right) = \sum_s \delta_s \left(u_s\left(\pi^{n'}\right) - u_s\left(\pi^n\right)\right) \geq 0.$$

Convexity in $\preceq$, thus, requires that the move from $\pi^n - \boldsymbol{\delta}$ to $\boldsymbol{\pi}^n$ be better news given the productive response to the dominating posterior (in terms of the order), $\boldsymbol{\pi}^{n'}$, than given the productive response for the dominated posterior, π^n. Therefore, the rank ordering of the vector

$$\mathbf{u}\left(\boldsymbol{\pi}^{n'}\right) - \mathbf{u}\left(\boldsymbol{\pi}^n\right),$$

is crucial to determining what type of $\boldsymbol{\delta}'s$ lead to more informative structures. An example illustrates. Suppose that $s' > s \Longrightarrow u_{s'}\left(\boldsymbol{\pi}^{n'}\right) - u_{s'}\left(\boldsymbol{\pi}^n\right) \geq u_s\left(\boldsymbol{\pi}^{n'}\right) - u_s\left(\boldsymbol{\pi}^n\right)$. Then,

$$\sum_{s=1}^{k} \delta_s \leq 0, \quad k = 1, 2, .., S-1$$

ensures convexity in $\preceq$ at $\boldsymbol{\delta}$ (recall $\sum_{s=1}^{S} \delta_s = 0$). More generally, W is convex in the order $\preceq$ at $\boldsymbol{\delta}'s$ that have their largest elements corresponding to the largest elements of $\mathbf{u}\left(\boldsymbol{\pi}^{n'}\right) - \mathbf{u}\left(\boldsymbol{\pi}^n\right)$.

When W is everywhere Gateaux differentiable then for an arbitrary order $\preceq$ and ordered pair $\boldsymbol{\pi}^o \preceq \boldsymbol{\pi}^*$:

$$W'\left(\boldsymbol{\pi}^{n'};\boldsymbol{\pi}^* - \boldsymbol{\pi}^o\right) - W'\left(\boldsymbol{\pi}^n;\boldsymbol{\pi}^* - \boldsymbol{\pi}^o\right) = \sum_s \left(\pi_s^* - \pi_s^o\right)\left(u_s\left(\pi^{n'}\right) - u_s\left(\pi^n\right)\right) \geq 0.$$

Hence, one will be able to rank posterior distributions according to $\pi^n \preceq_{\preceq} \pi^{n'}$ if productive responses are, in essence, always positively correlated with the difference between any two distributions ordered by $\preceq$. Thus, $\preceq_{\preceq}$ requires either placing some type of curvature restrictions upon U^* or restricting attention to those portions of U^* where those conditions apply. An example illustrates. Suppose we consider the ordering $\preceq_s$, then $\boldsymbol{\pi}^n \preceq_{\preceq_S} \boldsymbol{\pi}^{n'}$ will be satisfied if the rank ordering of $s' > s \Longrightarrow u_{s'}\left(\pi^{n'}\right) - u_{s'}\left(\pi^n\right) \geq u_s\left(\pi^{n'}\right) - u_s\left(\pi^n\right)$.

6 Concluding comments

The state-contingent production framework yields a natural characterization of the value of information in terms of convex sets. In particular, it is straightforward to derive upper and lower bounds for the value of information and rankings of the valuation for alternative problem settings and information structures.

The analysis presented above relies on the assumption of expected-utility maximization which implies that preferences are linear in probabilities. Decision-makers who do not satisfy the expected utility axioms have intrinsic preferences with respect to information (Grant, Polak, and Kajii 1988) and may therefore have negative *ex ante* willingness-to-pay for informative signals. The analysis could be extended to more general preferences in two ways. First, in some cases, quasi-concavity of preferences is sufficient to guarantee a positive value of information. Second, it may be possible to decompose the value of information into an intrinsic component, which may be positive or negative, and an instrumental component which measures the difference between the welfare associated with an optimal response to information and the welfare associated with a passive response. The latter component is necessarily positive.

Analysis of the value of information could also be applied to problems of contracting with asymmetric information. Quiggin and Chambers (1998) show how principal-agent problems involving moral hazard can be represented in a state-contingent framework. This framework could be extended to allow for the possibility that either the principal or the agent could acquire additional information.

Reference List

Athey, S. "Investment and Information Value for a Risk-Averse Firm." Working Paper, MIT, 2000.

———. "Monotone Comparative Statics Under Uncertainty." *Quarterly Journal of Economics* (forthcoming).

Athey, S., and J. Levin. "The Value of Information in Monotone Decision Problems." Working Paper, MIT, 2000.

Blackwell, D. "Comparison of Experiments." *Proceedings of the Second Berkeley Symposium on Mathematical Statistics* (1951): 93-102.

Chambers, R. G., and J. Quiggin. *Uncertainty, Production, Choice, and Agency: The State-Contingent Approach*. New York: Cambridge University Press, 2000.

Färe, R. *Fundamentals of Production Theory*. Berlin: Springer-Verlag, 1988.

Grant, S., A. Kajii, and B. Polak. "Intrinsic Preferences for Information." *Journal of Economic Theory* 83 (1998): 233-59.

Lehmann, E. L. "Comparing Location Experiments." *Annals of Statistics* 16 (1988): 521-33.

Marshall, A., and I. Olkin. *Inequalities: Theory of Majorization and Its Applications*. New York: Academic Press, 1979.

McFadden, D. "Cost, Revenue, and Profit Functions." *Production Economics: A Dual Approach to Theory and Applications*. eds. M. Fuss, and D.McFaddenAmsterdam: North-Holland, 1978.

Milgrom, P., and C. Shannon. "Monotone Comparative Statics." *Econometrica* 62 (1994): 157-80.

Newbery, D., and J. Stiglitz. *The Theory of Commodity Price Stabilisation--A Study in the Economics of Risk*. Oxford: Oxford University Press, 1981.

Ormiston, M. Schlee E. "Comparative Statics Under Uncertainty for a Class of Economic Agents." *Journal of Economic Theory* 61 (1993): 412-22.

Ormiston, M., and E. Schlee. "Necessary Conditions for Comparative Statics Under Uncertainty." *Economics Letters* 40 (1992): 429-34.

Quiggin, J., and R. G. Chambers. "Risk Premiums and Benefit Measures for Generalized Expected Utility Theories." *Journal of Risk and Uncertainty* 17 (1998): 121-38.

Rockafellar, R. T. *Convex Analysis*. Princeton: Princeton University Press, 1970.

Sandmo, A. "On the Theory of the Competitive Firm Under Price Uncertainty." *American Economic Review* 61 (1971): 65-73.

Topkis, D. M. *Supermodularity and Complementarity*. Princeton: Princeton University Press, 1998.