Another Look at the Radner-Stiglitz Nonconcavity in the Value of Information*

Hector Chade
Department of Economics
Arizona State University

Edward Schlee
Department of Economics
Arizona State University

April 2000

Abstract

This paper revisits the well-known result of Radner and Stiglitz (1984) which shows that, under certain conditions, the value of information exhibits increasing marginal returns over some range. Their result assumes that both the number of states and the number of signal realizations are finite, assumptions which preclude most analyses of optimal information acquisition. We provide a set of sufficient conditions that yields this 'nonconcavity' in the value of information in a general framework; the role that these conditions play is clarified and illustrated with several examples. We also discuss the robustness of the nonconcavity result, and the difficulties involved in getting the value of information to be globally concave.

*We would like to thank Alejandro Manelli, Lones Smith, Peter Sørensen, Len Mirman, Eric Balder, and Nicholas Yannelis for their helpful comments, and seminar participants at the University of Illinois, the University of Arizona, the Brown Bag Lunch Workshop at Arizona State University, and at the Southern California Economic Theory Conference at Santa Barbara.

1 Introduction

Is there an intrinsic nonconcavity to the value of information? In an important and influential paper, the analysis of Radner and Stiglitz (1984) suggests that there is. They demonstrated, in a seemingly general model, that the marginal value of a small amount of information is zero. Since the marginal value of (costless) information is always nonnegative, this finding implies that, unless information is useless and hence always of zero value, it must exhibit increasing marginal returns over some range. Radner and Stiglitz do present a few examples that violate their assumptions for which information exhibits decreasing marginal returns, so it is clear that the value of information does not *always* exhibit a nonconcavity. Yet, the conditions under which they obtain the nonconcavity seem at first glance to be fairly innocuous. They assume that the number both of states and signal realizations are finite. They index the information structure, represented by a Markov matrix of state-conditional signal distributions, by a parameter taking on values in the unit interval, with a zero value of the parameter representing null information. They then impose two assumptions: one is that this Markov matrix is a differentiable function of the index parameter at the zero value; the other is a continuity restriction on a particular selection from the correspondence of maximizers. These conditions appear to be standard smoothness and continuity assumptions: although it may not always hold, their result cannot easily be dismissed as depending on exotic assumptions.

As Radner and Stiglitz note, this nonconcavity has several important implications: the demand for information will be a discontinuous function of its price (under linear pricing); agents will not buy 'small' quantities of information; and agents will tend to specialize in information production. As with any nonconcavity, it will generally tend to complicate any analysis of information acquisition, and it can also have important consequences in applications. For example, it may preclude the existence of a competitive equilibrium (Wilson (1975), Radner (1989)), or the existence of a linear Rational Expectations Equilibrium (Laffont (1985)) if information can be acquired by agents; or it may have substantial effects on the organization of production when moral hazard is present and there is a demand for monitoring (Singh (1985)).[1]

[1] See also Arrow (1985) and Bradford and Kelejian (1977) for additional remarks on the effects of the nonconcavity in the value of information.

The nonconcavity has proven to be especially vexing to the literature on active learning or experimentation.[2] In this literature, an agent takes an action in each period in the face of uncertainty about a parameter that affects her payoff. The agent observes a random signal that depends both on her action and the value of the unknown parameter; after observing the signal, she updates beliefs and then chooses again. An interesting feature of such models is that the agent not only learns, but can affect how much she learns by varying her actions: she sacrifices utility today to increase information available tomorrow. As an example, a price setting firm may face an uncertain demand function: by varying the price, the firm may be able to affect how much it learns about its demand. Thus the present action acts as an index of the informativeness of the 'experiment' that the agent observes. If the value of information is not concave in the present action, then the analysis of optimal experimentation is made much more complex. We will show later that our sufficient conditions for the nonconcavity to hold are met by a broad class of experimentation problems commonly found in the literature. Moreover, some recent papers have considered experimentation in strategic settings, specifically, an industry of firms that learn about demand while competing with one another (Harrington (1995); Mirman, Samuelson and Schlee (1994)). In these models, the nonconcavity means that the first period best reply mappings of firms may not be convex-valued, so that pure strategy equilibria may not exist.[3] Mirman, Samuelson and Schlee (1994) thus examine mixed strategy equilibria, while Harrington (1995) restricts attention to models in which firms discount the future heavily. Neither approach is very satisfactory in an oligopoly model.

Besides complicating models of information acquisition, a more fundamental question remains: why is it that information should intrinsically exhibit increasing marginal returns (starting from no information)? While information as a commodity is admittedly special, it is still somewhat puzzling that it cannot generally exihibit diminishing marginal returns.

The purpose of this paper is to re-examine the conditions under which the Radner-Stiglitz nonconcavity holds, that is, a small amount of information

[2]Among others, see Mirman, Samuelson and Urbano (1993), Tonks (1984), Kihlstrom, Mirman and Postlewaite (1984), Harrington (1995), Easley and Kiefer (1988), and Mirman, Samuelson and Schlee (1994).

[3]Other sufficient conditions for the existence of pure strategy equilibria are not helpful here either. For example, the restriction to supermodular games would eliminate many of the examples in Mirman, Samuelson and Schlee (1994).

has zero marginal value. (Of course, even if the marginal value of a small amount of information is positive, the value of information can still exhibit a nonconcavity; we shall refer to the Radner-Stiglitz nonconcavity as that which arises from a zero marginal value at null information). We propose to do so in a fairly general Bayesian decision problem. All of the aforementioned applications have assumed either an infinite number of signal realizations or an infinite number of states, unlike the original Radner-Stiglitz framework. One of our main goals is to extend their theorem to this more general model. We provide a set of sufficient conditions under which the Radner-Stiglitz nonconcavity arises, and show that it subsumes the Radner-Stiglitz framework as a special case, as well as most of the applications encountered in the literature where the nonconcavity is present. Although some of the assumptions are purely technical, most are substantive: we present examples showing that their failure leads to a failure of the nonconcavity.

Besides extending their theorem, we also want to clarify the role that these conditions play. Of particular interest is the assumption by Radner and Stiglitz of the existence of a particular continuous selection from the correspondence of maximizers (one that is constant in the signal realization at null information, our assumption **A0** below). We provide sufficient conditions separately on the information structure and the decision maker's utility function and prior beliefs to ensure the existence of such a selection. Several examples are provided to illustrate the role played by the assumptions imposed.

We also argue that the more general setting presented here will help us to evaluate the robustness of the nonconcavity. Since there are important contributions to the literature on information acquisition that do not exhibit the nonconcavity,[4] we revisit some of them and discuss why these information structures avoid the problem in the models analyzed in those contributions. We also present examples to illustrate the difficulties involved in proving results to get the value of information to be globally *concave*. The tentative conclusion we draw from our Theorems and examples is that, although our sufficient conditions for the Radner-Stiglitz nonconcavity are strong, and one can construct specific models that yield a concave value of information, a nonconcavity in the value of information seems difficult to rule out in a model of much generality. Whether the reader agrees with this interpretation of our

[4]In particular, Kihlstrom (1974), Freixas and Kihlstrom (1984), Arrow (1985), Moscarini and Smith (1999, 2000).

results, our hope is to stimulate thinking on the appropriate 'functional form restrictions' to impose on the information structure of a problem.

Moscarini and Smith (2000) hold out some hope that the nonconcavity might at least be tamed, if not vanquished. Assume that both the number of actions and states are finite. They show that, if the quantity of information is measured by the number of (state-conditionally) independent observations from an experiment, then the marginal value of information *eventually* falls as the number of observations increases. Hence, if the price of observations is low enough, the demand for information will be well-behaved. By focusing on the large demand case and measuring information by the sample size, they thereby avoid the 'small quantity' problem emphasized by the Radner and Stiglitz result. We return to their paper after presenting our main results.

The paper is organized as follows. Section 2 presents a detailed description of the general decision problem we consider. In Section 3, we state the Radner and Stiglitz theorem and provide an intuitive explanation of its main assumptions. The main results are derived in Section 4, where we prove a general theorem and illustrate the role of the assumptions with examples and corollaries covering some special cases that are often assumed in the literature. Section 5 uses the main results to discuss some important contributions on the demand for information that do not exhibit the nonconcavity result, casting some light on the reasons for its absence. Section 6 concludes.

2 The Model

A Bayesian agent who is uncertain about the state of the world must choose an action after observing the realization of a random variable that is possibly correlated with the state. We index the set of information structures by a real parameter $\theta \in \Theta$.

The formal description of the model is the following:

- The set of states of the world S is a complete, separable metric space, endowed with the Borel σ-algebra $\mathcal{B}_S$; the measure $\mu : \mathcal{B}_S \to [0,1]$ represents the prior beliefs of the decision maker.
- The set of signals the decision maker can observe is denoted by Y, and it is a complete, separable metric space with Borel σ-algebra $\mathcal{B}_Y$.
- $\Theta = [0,1]$ is the index set.

- For each θ, $Q(\cdot \mid \cdot, \theta)$ is a stochastic kernel on Y given S that represents an information structure available to the agent; i.e, for each $s \in S$, $Q(\cdot \mid s, \theta) : \mathcal{B}_Y \to [0,1]$ is a probability measure, and for each $C \in \mathcal{B}_Y$, $Q(C \mid \cdot, \theta) : S \to [0,1]$ is $\mathcal{B}_S$-measurable. Different values of θ correspond to different information structures. An uninformative information structure is represented by $\theta = 0$; formally, for all $s, s' \in S$, $Q(\cdot \mid s, 0) = Q(\cdot \mid s', 0)$.

- The action space A is a complete, separable metric space with Borel σ-algebra $\mathcal{B}_A$.

- $u : A \times S \longrightarrow \mathbb{R}$ is the decision maker's vonNeumann-Morgenstern utility function; it is assumed to be jointly continuous and bounded.

- $\mathcal{D}$ is the set of all functions $d : Y \to A$ that are $(\mathcal{B}_Y, \mathcal{B}_A)$-measurable (i.e., $d^{-1}(B) \in \mathcal{B}_Y$ for each $B \in \mathcal{B}_A$). The set $\mathcal{D}$ contains the decision functions or strategies available to the decision maker. [5]

Since the agent can condition her decision on the signal observed, her problem is:

$$\max_{d \in \mathcal{D}} \int_S \int_Y u(d(y), s) Q(dy \mid s, \theta) \mu(ds). \tag{1}$$

Let $V(\theta)$ be the value function of the problem, which is interpreted as the value of the information structure θ, and let $D^*(\theta)$ be the correspondence of maximizers. That is,

$$V(\theta) = \max_{d \in \mathcal{D}} \int_S \int_Y u(d(y), s) Q(dy \mid s, \theta) \mu(ds),$$

and

$$D^*(\theta) = \{d \in \mathcal{D} : \int_S \int_Y u(d(y), s) Q(dy \mid s, \theta) \mu(ds) \ = V(\theta)\}.$$

A selection from this correspondence will be denoted by $d^*(y, \theta)$, in order to emphasize the dependence on θ. Given a selection $d^*(y, \theta)$, then

$$V(\theta) = \int_S \int_Y u(d^*(y, \theta), s) Q(dy \mid s, \theta) \mu(ds). \tag{2}$$

[5] Most of the results of the paper also hold in the slightly more general case where the set $\mathcal{D}$ depends on θ, say $\mathcal{D}(\theta)$, and $\mathcal{D}(\theta_1) \subseteq \mathcal{D}(\theta_2)$ whenever $\theta_1 > \theta_2$. The only exception is the existence of the right-hand derivative of the value function.

Although we shall prove the main results using the *normal form* of the problem described by (1), an alternative and common way to analyze this decision problem is in its *extensive form*, which exploits the sequential structure of the model more explicitly.

The extensive form is derived as follows. Fix $\theta \in [0,1]$; given $Q(\cdot \mid \cdot, \theta)$ and $\mu(\cdot)$, then by Theorem 8.5 in Stokey, Lucas, and Prescott (1989) there exists a unique probability measure $\pi(\cdot \mid \theta)$ on $(S \times Y, \mathcal{B}_S \times \mathcal{B}_Y)$ such that, for each $B \in \mathcal{B}_S$ and $C \in \mathcal{B}_Y$,

$$\pi(B \times C \mid \theta) = \int_B Q(C \mid s, \theta)\mu(ds).$$

Let $\nu(C \mid \theta) = \pi(S \times C \mid \theta)$ be the marginal of $\pi(\cdot \mid \theta)$ on $\mathcal{B}_Y$; i.e., $\nu(C \mid \theta)$ is the probability that $y \in C$ if the information structure is θ. Under the assumptions made on S and Y, given $\pi(\cdot \mid \theta)$ and $\nu(\cdot \mid \theta)$, then by Corollary 7.27.2 in Bertsekas and Shreve (1978) there exists a stochastic kernel $P(\cdot \mid \cdot, \theta)$ on S given Y such that

$$\pi(B \times C \mid \theta) = \int_C P(B \mid y, \theta)\nu(dy \mid \theta).$$

The stochastic kernel $P(\cdot \mid y, \theta)$ can be interpreted as a version of the posterior beliefs of the decision maker after observing y, when the information structure is θ; for each y, her problem is then

$$\max_{a \in A} \int_S u(a, s)P(ds \mid y, \theta). \tag{3}$$

Let $U(y, \theta)$ and $A^*(y, \theta)$ be the value function and the correspondence of maximizers of problem (3); given any $(\mathcal{B}_Y, \mathcal{B}_A)$-measurable selection $d^*(y, \theta)$ from this correspondence, then the value function in the extensive form representation is

$$\begin{aligned} V(\theta) &= \int_Y \int_S u(d^*(y, \theta), s)P(ds \mid y, \theta)\nu(dy \mid \theta) \\ &= \int_Y U(y, \theta)\nu(dy \mid \theta) \qquad (4) \end{aligned}$$

The extensive form representation affords a simple proof of existence of a solution and satisfaction of the measurability requirements implicit in (4).

Proposition 1 *If A is a compact metric space then, for each $\theta \in \Theta$,*

(i) $U(y, \theta)$ is $\mathcal{B}_Y$-measurable and bounded;

(ii) $A^*(y, \theta)$ is nonempty and admits a $(\mathcal{B}_Y, \mathcal{B}_A)$-measurable selection $d^*(y, \theta)$.

Proof: Fix θ and set $g(a, y, \theta) = \int_S u(a, s) P(ds \mid y, \theta)$; continuity of $u : A \times S \to \mathbb{R}$ and $\mathcal{B}_Y$-measurability of the stochastic kernel imply that $g(\cdot, y, \theta) : A \to \mathbb{R}$ is continuous and $g(a, \cdot, \theta) : Y \to \mathbb{R}$ is $\mathcal{B}_Y$-measurable. Since A is compact, (i) and (ii) follow from the Measurable Maximum Theorem (Aliprantis and Border (1999), Theorem 17.18). This completes the proof of the proposition. ■

Until section 4.2, we will impose the following assumption on the correspondence of maximizers:

A0: There exists a $(\mathcal{B}_Y, \mathcal{B}_A)$-measurable selection $d^*(y, \theta)$ with the following properties: (i) $\lim_{\theta \to 0+} d^*(y, \theta) = d^*(y, 0)$ for every y, (ii) $d^*(y, 0) = a_0^*$ for every y.

Radner and Stiglitz (1984) used this assumption in their model with a finite number of states and signals to derive their nonconcavity result. In words, **A0** says that there exists an optimal decision that is 'continuous in θ and flat in y' at $\theta = 0$. Since this imposes conditions jointly on the information structure *and* the decision maker's utility function and prior beliefs, it is not entirely satisfactory. One of our goals will be to justify **A0** from conditions imposed separately on those elements, and explain their roles in yielding the conclusion.

3 The Question

We are now ready to formulate our main question: under what conditions is the marginal value of a small amount of information equal to zero? More precisely, we seek conditions on the information structure such that, along with **A0**, imply that $V'(0+)$ exists and equals zero. Note that the value function $V : \Theta \to \mathbb{R}$ excludes any cost of information acquisition. Since, in the abstract, there may be no obvious natural units to measure the amount of information, we should stress that this question has meaning only in the context of a broader decision problem that involves *choosing* an information structure; and very often in such problems there is a natural index of the available information structures. A simple, but very useful, formulation is

$$\max_{\theta \in [0,1]} V(\theta) - C(\theta) \tag{5}$$

Here $C : \Theta \to \mathbb{R}$ represents the cost of different information structures. As an example, the decision maker could be a firm that is uncertain about its market demand. The parameter θ could represent the number of hours spent on marketing research, with zero hours yielding no information; $V(\theta)$ then is the maximum expected profit from operating in a market when the firm spends θ hours doing market research at a cost of $C(\theta)$ dollars. More generally, all standard two-period experimentation models can be written in form (5). In such models, an agent takes some action in the first period; a noisy signal of the state is then revealed; the agent updates beliefs and then chooses an action in the second period. In our notation, the utility function $u(a, s)$ gives the second period utility from taking action a (an element of $[0, 1]$ say) under state s. The parameter θ represents a first period action that affects the distribution of the observed signal, and hence how much information the agent has in the second period; the value function $V(\theta)$ then gives the maximum second period expected utility as a function of the first period action θ. Finally, the cost function equals the difference between first period expected utility from choosing θ under the prior belief and the maximum first period utility; formally (assuming that u is the utility function in both period 1 and period 2),

$$C(\theta) = \overline{V} - \int_S u(\theta, s)\mu(ds),$$

where

$$\overline{V} = \max_{\theta' \in [0,1]} \int_S u(\theta', s)\mu(ds).$$

The prototypical problem studied in the optimal experimentation literature is that of a firm learning about demand.[6] That this fits (5) is illustrated by the following example:

[6] See, among others, McLennan (1984), Mirman, Samuelson, and Urbano (93), Grossman, Kihlstrom, and Mirman (1977), Aghion et al. (1991), Treffler (1993), and Creane (1994).

Example 1: Let the demand function be given by $f(p, s, \varepsilon)$, where p is the market price, s the state of demand, and ε the realization of a (i.i.d.) random noise variable. The firm chooses price at date 1, observes the sales realization (but neither s nor ε), updates beliefs about s, and then chooses a date 2 price. For concreteness, specialize further to the case of linear demand: $f(p, s, \varepsilon) = (\alpha - p)s + \varepsilon$. In terms of our notation $u(a, s) = ((\alpha - a)s + E[\varepsilon])\,(a - k)$, where k is a constant marginal cost, a is the second period price, and $E[\varepsilon]$ is the expected volume of noise demand; $u(\cdot)$ gives the date 2 profit as a function of the date 2 price and the demand parameter s. Notice that a first period price of α gives no information about the permanent part of demand (since only 'noise traders' buy at this price). Thus we can define $\theta = \alpha - p$, so that $\theta = 0$ corresponds to a null information structure. The firm sets θ in period one, observes sales of $\theta s + \varepsilon$, updates beliefs about s, and then sets the date 2 price. In this case the cost function $C(\cdot)$ describes the expected profit loss as a result of deviating from the myopically optimal price, and $V(\theta)$ gives the maximum second period profit from charging a price of $\alpha - \theta$ at date 1.

Now, as long as the cost function $C(\cdot)$ is increasing with $C'(0+) > 0$, the objective function in (5) will not be concave if $V'(0+) = 0$ and information has positive value for some $\theta > 0$. From the perspective of these applications, we can rephrase our question as determining whether the objective function in (5) can be concave for a cost function with positive marginal cost at $\theta = 0$. As we have noted, Radner and Stiglitz (1984) answered this question for the special case in which the set of signal realizations Y and the set of states S are both finite; that is, $Q(C \mid s, \theta) = \sum_{y \in C} q(y \mid s, \theta)$ for each $C \subseteq Y$, where $q(y \mid s, \theta)$ is the probability of observing y if the state is s and the information structure is θ. They showed the following result:[7]

Proposition 2 *Assume that*
a) A0 holds;
b) $q(y \mid s, \theta)$ is differentiable with respect to θ at $\theta = 0$.
Then $D^+V(0) = \limsup_{\theta \to 0+} \frac{V(\theta) - V(0)}{\theta} \leq 0$.

In other words, if $V'(0+)$ exists, it must be nonpositive. Condition b) ensures that the information structure varies smoothly with θ. As we have

[7] They actually proved the result in a slightly more general set up that is a special case of the framework described in footnote 5.

noted, **A0** jointly restricts the information structure and the decision maker's utility function and prior beliefs. Proposition 3 (in Section 4.2) gives conditions on the primitives of the problem that ensure that **A0** holds in the general model. Corollary 4 deals with the finite case considered here: in particular, it asserts that **A0** holds if i) $u(\cdot, s) : A \to \mathbb{R}$ is strictly concave (assuming that A is convex); and ii) if $q(y \mid s, 0) > 0$ for all $y \in Y$. The first ensures that there is a unique optimal action for each posterior belief in the *extensive* form of the problem; and the second says that the null information structure has full support on the signals. Intuitively, it is easy to see how the conclusion can fail if i) doesn't hold. At $\theta = 0$, the posterior belief of course equals the prior belief for all signal realizations; for $\theta > 0$, the posterior will differ from the prior for some values of y if the experiment is informative. If there is more than one optimal action at the prior, then even 'small' changes in the posterior can result in 'large' changes in the set of optimal actions.[8] Hence a small amount of information can have a positive marginal value. To understand the role of ii), recall that the posterior belief that the state is s after observing y for information structure indexed by θ is given by: $P(\{s\} \mid y, \theta) = \frac{\mu_s q(y|s,\theta)}{\sum_{s \in S} \mu_s q(y|s,\theta)}$, where $\mu_s = \mu(\{s\})$ for all $s \in S$. The smoothness assumption b) is not sufficient to ensure even the continuity of $P(\{s\} \mid y, \theta)$ in θ at $\theta = 0$, as the following example illustrates:

Example 2: Let $S = \{s_L, s_H\}$, $\mu_{s_L} = \mu_{s_H} = \frac{1}{2}$, $Y = \{y_1, y_2\}$ and, for all $\theta \in [0, 1]$, $q(y_2 \mid s_H, \theta) = 1$ and $q(y_1 \mid s_L, \theta) = \theta$. Then $P(\{s_H\} \mid y_1, \theta) = 0$ for all $\theta > 0$, so that $P(\{s_H\} \mid y_1, \theta)$ doesn't converge to the prior of $\frac{1}{2}$ as θ tends to 0. Without continuity of the posterior in θ at $\theta = 0$, small changes in θ produce large changes in the distribution of posterior beliefs, which may have a positive marginal value.

Assuming ii) holds, condition b) then ensures that the posterior belief will also be *differentiable* at $\theta = 0$. Since the posterior never strays far from the prior as θ increases from zero, a small increase in θ cannot improve information very much over null information, and hence under i) such a change has a zero marginal value at $\theta = 0$. In other words, under the smoothness assumption, a small increase in θ from $\theta = 0$ has only a second order effect on expected utility.

[8]The Maximum Theorem merely ensures upper hemicontinuity of the correspondence of maximizers. Example 7 shows that the nonconcavity can fail if the optimal choice in the extensive form (3) is not single-valued.

As we noted in the introduction our interest in this question is twofold: first, we think it is intrinsically interesting to determine whether information 'generally' exhibits increasing marginal returns; and second, the nonconcavity complicates the analysis of decision problems that involve information production or acquisition, and can have important consequences in applications. Now, the differentiability condition in b) looks like a mild regularity requirement; hence, it may appear that the nonconcavity is a fairly robust phenomenon. The natural setting of problems of the form (5), however, is one either with an infinite number of signals or states, or at least that is the setting in which most of the analysis of information acquisition has been carried out. Thus we seek to determine conditions under which the conclusion of Proposition 2 extends to the more general set up described in Section 2, and to identify some important specifications for which the nonconcavity fails.

4 Main Results

The purpose of this section is to investigate whether, given a selection with the properties stated in **A0**, the value of the information structure θ

$$V(\theta) = \int_S \int_Y u(d^*(y,\theta), s) Q(dy \mid s, \theta) \mu(ds)$$

can be concave in θ. It will be shown that, under certain conditions, if we start with an uninformative information structure, then the marginal returns of a small amount of information is zero; this in turn implies that, under those conditions, $V(\theta)$ cannot be globally concave.

An informal way to explain what we do below is the following: suppose for a moment that $d^*(y,\theta)$ and $Q(\cdot \mid s, \theta)$ are differentiable in θ, and assume that we can 'pass' the derivative through the integral. Then the Envelope Theorem implies that:

$$V'(\theta) = \int_S \int_Y u(d^*(y,\theta), s) Q_\theta(dy \mid s, \theta) \mu(ds).$$

Evaluating this at $\theta = 0$ and using **A0** yields

$$\begin{aligned} V'(0) &= \int_S \int_Y u(a_0^*, s) Q_\theta(dy \mid s, 0)\mu(ds) \\ &= \int_S u(a_0^*, s)\left(\int_Y Q_\theta(dy \mid s, 0)\right)\mu(ds) \\ &= 0, \end{aligned} \tag{6}$$

where the last step follows from the fact that $\int_Y Q(dy \mid s, \theta) = 1$, and therefore $\int_Y Q_\theta(dy \mid s, 0) = 0$. In other words, the marginal value of a small amount of information is zero if we start with an uninformative information structure. To be sure, this result was derived under the strong assumption of differentiability of the selection and without justifying the interchange of the derivative and the integral; moreover, we did not explain the meaning of the 'derivative' of the stochastic kernel, and whether integration with respect to Q_θ was well-defined. The results below provide sets of sufficient conditions that i) relax the differentiability assumption on the selection; ii) justify the interchange of differentiability and integration; and iii) gives a meaningful interpretation to Q_θ, so that $V'(0+)$ exists and equals zero.

4.1 Generalization of the Radner-Stiglitz Theorem

We now present a set of sufficient conditions for the nonconcavity result in our general Bayesian decision framework. It will be shown that the finite case and all the applications cited in the introduction that exhibit the nonconcavity are subsumed as special cases of our theorem.

Let $ca(\mathcal{B}_Y)$ be the space of finite signed measures on $(Y, \mathcal{B}_Y)$, and endow it with the total variation norm $\|\lambda\| = \mid \lambda \mid (Y)$ (Halmos (1950), pp.122-123).

We will impose the following 'smoothness' assumption, which generalizes the differentiability condition *b*) of Proposition 2 for the finite case.

A1: For each $s \in S$ and $C \in \mathcal{B}_Y$,

$$\lim_{\theta \to 0+} \frac{Q(C \mid s, \theta) - Q(C \mid s, 0)}{\theta} = Q_\theta(C \mid s, 0)$$

exists in $\mathbb{R}$.

By Corollary 4 in Dunford and Schwartz (1957, p.160) $Q_\theta(\cdot \mid s, 0)$ is an element of $ca(\mathcal{B}_Y)$. It turns out that **A1** is not sufficient for our purposes (see Example 4 below); we shall also require that the convergence condition in **A1** hold in the total variation norm.

A2: For each $s \in S$,

$$\lim_{\theta \to 0+} \left\| \frac{Q(\cdot \mid s, \theta) - Q(\cdot \mid s, 0)}{\theta} - Q_\theta(\cdot \mid s, 0) \right\| = 0.$$

The proof of the Theorem uses the following result, whose proof is in the appendix:

Lemma 1 *Let $(X, \mathcal{F})$ be a measurable space, $ca(\mathcal{F})$ the space of finite signed measures on $\mathcal{F}$ endowed with the total variation norm, $\{\nu_n\}$ a sequence in $ca(\mathcal{F})$ that converges in the total variation norm to ν, and $\{f_n\}$ a sequence of bounded measurable functions that converge pointwise to f. Then*

$$lim_{n \to \infty} \int f_n d\nu_n = \int f d\nu.$$

Theorem 1 *Assume that*

a) A0, A1, and A2 hold;

b) There exists a μ-integrable function $M : S \to \mathbb{R}$ such that, for every $\theta \in (0, 1]$ and $s \in S$,

$$\left\| \frac{Q(\cdot \mid s, \theta) - Q(\cdot \mid s, 0)}{\theta} \right\| \leq M(s).$$

Then $V'(0+)$ exists and it is equal to zero.

Proof: We first show that $\limsup_{\theta \to 0+} \frac{V(\theta) - V(0)}{\theta} \leq 0$. As in Radner and Stiglitz (1984), write $\frac{V(\theta)-V(0)}{\theta} = \frac{T_1(\theta)}{\theta} + \frac{T_2(\theta)}{\theta}$, where

$$\begin{aligned} T_1(\theta) &= \int_S \int_Y u(d^*(y, \theta), s) Q(dy \mid s, \theta) \mu(ds) \\ &\quad - \int_S \int_Y u(d^*(y, \theta), s) Q(dy \mid s, 0) \mu(ds), \\ T_2(\theta) &= \int_S \int_Y u(d^*(y, \theta), s) Q(dy \mid s, 0) \mu(ds) \\ &\quad - \int_S \int_Y u(a_0^*, s) Q(dy \mid s, 0) \mu(ds). \end{aligned}$$

Since a_0^* is optimal at $\theta = 0$, it follows that $T_2(\theta) \leq 0$ and $\limsup_{\theta \to 0+} \frac{T_2(\theta)}{\theta} \leq 0$.

Consider

$$\frac{T_1(\theta)}{\theta} = \int_S \int_Y u(d^*(y, \theta), s)(\frac{Q(dy \mid s, \theta) - Q(dy \mid s, 0)}{\theta})\mu(ds).$$

Assumption **A1** ensures that this integral is well-defined for every θ.[9] Moreover, since

$$\lim_{\theta \to 0+} \frac{Q(C \mid s, \theta) - Q(C \mid s, 0)}{\theta} = Q_\theta(C \mid s, 0)$$

for each measurable set C, then $Q_\theta(\cdot \mid s, 0) \in ca(\mathcal{B}_Y)$ and, being the pointwise limit of $\mathcal{B}_S$-measurable functions, it is $\mathcal{B}_S$-measurable.

We now prove that

$$\lim_{\theta \to 0+} \frac{T_1(\theta)}{\theta} = \int_S \int_Y u(a_0^*, s) Q_\theta(dy \mid s, 0)\mu(ds). \tag{7}$$

Take any sequence θ_n converging to 0, and let

$$\begin{aligned} h_n(s) &= \int_Y u(d^*(y, \theta_n), s)(\frac{Q(dy \mid s, \theta_n) - Q(dy \mid s, 0)}{\theta_n}), \\ h(s) &= \int_Y u(a_0^*, s) Q_\theta(dy \mid s, 0). \end{aligned}$$

Given **A0**, **A1**, and **A2**, it follows from Lemma 1 that $h_n(s)$ converges pointwise to $h(s)$; moreover, condition $b)$ ensures that the convergence is dominated. For

$$\begin{aligned} \mid h_n(s) \mid &= \mid \int_Y u(d^*(y, \theta_n), s)(\frac{Q(dy \mid s, \theta_n) - Q(dy \mid s, 0)}{\theta_n}) \mid \\ &\leq B \| \frac{Q(\cdot \mid s, \theta_n) - Q(\cdot \mid s, 0)}{\theta_n} \| \\ &\leq BM(s), \end{aligned}$$

[9] This follows from Stokey, Lucas, and Prescott (1989), Theorem 8.4 and its Corollary, which also holds for signed kernels like the ones considered here (just decompose the signed kernel into its positive and negative variation).

where $B < \infty$ is such that $\mid u(a,s) \mid \leq B$ (Royden (1988), p.275). It follows by the Lebesgue Dominated Convergence Theorem (LDCT) that

$$\lim_{n\to\infty} \int_S h_n(s)\mu(ds) = \int_S h(s)\mu(ds),$$

and, since $\{\theta_n\}$ was an arbitrary sequence converging to zero, (7) holds.

If we could show that $\int_Y Q_\theta(dy \mid s,0) = 0$, it would then follow by (6) that

$$\lim_{\theta\to 0+} \frac{T_1(\theta)}{\theta} = 0$$

But

$$\int_Y \frac{Q(dy \mid s,\theta) - Q(dy \mid s,0)}{\theta} = \frac{\int_Y Q(dy \mid s,\theta) - \int_Y Q(dy \mid s,0)}{\theta} = 0.$$

Therefore, another application of Lemma 1 yields[10]

$$0 = \lim_{\theta\to 0+} \int_Y \frac{Q(dy \mid s,\theta) - Q(dy \mid s,0)}{\theta} = \int_Y Q_\theta(dy \mid s,0).$$

Hence,

$$\limsup_{\theta\to 0+} \frac{V(\theta) - V(0)}{\theta} \leq \limsup_{\theta\to 0+} \frac{T_1(\theta)}{\theta} + \limsup_{\theta\to 0+} \frac{T_2(\theta)}{\theta} \leq 0.$$

Finally, since $V(\theta) - V(0) \geq 0$, we have that $\liminf_{\theta\to 0+} \frac{V(\theta)-V(0)}{\theta} \geq 0$. Therefore $V'(0+)$ exists and it is equal to zero. This completes the proof. ■

The verification of assumptions **A1** and **A2** may be a nontrivial task. We now turn to consider some special cases commonly found in applications, and show that they satisfy **A1** and **A2**.

Consider first the case in which there is a σ-finite measure $\nu : \mathcal{B}_Y \to [0,\infty]$ such that, for each $(s,\theta) \in S \times \Theta$, $Q(\cdot \mid s,\theta)$ has a density $q(y \mid s,\theta)$ with respect to ν, that is also μ-integrable. That is, for every $C \in \mathcal{B}_Y$ and $(s,\theta) \in S \times \Theta$,

$$Q(C \mid s,\theta) = \int_C q(y \mid s,\theta)\nu(dy).$$

[10] Set $f_n = 1$ for every n.

In particular, this includes the important case in applications where Y is a (Borel) subset of $\mathbb{R}^n$ (endowed with the Euclidean metric), ν is the n-dimensional Lebesgue measure, and $q(\cdot \mid s, \theta)$ is one of the familiar density functions defined on $\mathbb{R}^n$. It also covers the case where, for every $\theta \in \Theta$, the stochastic kernels are mutually absolutely continuous, and then ν is just $Q(\cdot \mid s, 0)$. Finally, it includes the countable case if ν is the counting measure defined on the σ-field of all subsets of Y (endowed with the discrete metric); it is straightforward to check that $Q(\cdot \mid s, \theta)$ is absolutely continuous with respect to ν for each (s, θ). Obviously, this subsumes the finite case considered by Radner and Stiglitz.

The value of information structure θ in this case becomes

$$V(\theta) = \int_S \int_Y u(d^*(y, \theta), s) q(y \mid s, \theta) \nu(dy) \mu(ds).$$

Corollary 1 (Absolutely Continuous Stochastic Kernel) *Assume that*

a) A0 holds;

b) $q(y \mid s, \theta)$ is differentiable with respect to θ at $\theta = 0$, and there is a $\nu \times \mu$-integrable function $z(y, s)$ such that $\mid \frac{q(y|s,\theta) - q(y|s,0)}{\theta} \mid \leq z(y, s)$ for every (y, s) and $\theta \in (0, 1]$.

Then $V'(0+)$ exists and it is equal to zero.

Proof: It is enough to show that **A1**, **A2**, and the integrability condition b) of Theorem 1 are satisfied. Given any set $C \in \mathcal{B}_Y$, then for any $s \in S$ and $\theta \in (0, 1]$,

$$\frac{Q(C \mid s, \theta) - Q(C \mid s, 0)}{\theta} = \int_C \frac{q(y \mid s, \theta) - q(y \mid s, 0)}{\theta} \nu(dy).$$

Since the integrand is dominated, the LDCT implies

$$\lim_{\theta \to 0+} \frac{Q(C \mid s, \theta) - Q(C \mid s, 0)}{\theta} = \int_C q_\theta(y \mid s, 0) \nu(dy).$$

Define

$$Q_\theta(C \mid s, 0) = \int_C q_\theta(y \mid s, 0) \nu(dy).$$

Since $q_\theta(y \mid s, 0)$ is integrable, then $Q_\theta(\cdot \mid s, 0)$ is a finite signed measure for each $s \in S$. This shows that **A1** holds.

Consider now **A2**. The difference between $\frac{Q(\cdot|s,\theta)-Q(\cdot|s,0)}{\theta}$ and $Q_\theta(\cdot \mid s, 0)$ is a finite signed measure for each $s \in S$ and $\theta \in (0, 1]$, and it can be written as

$$\lambda(C \mid s,\theta) = \int_C f(y \mid s,\theta)\nu(dy)$$

for every $C \in \mathcal{B}_Y$, where

$$\lambda(\cdot \mid s,\theta) = \frac{Q(\cdot \mid s,\theta) - Q(\cdot \mid s,0)}{\theta} - Q_\theta(\cdot \mid s,0),$$

and

$$f(y \mid s,\theta) = \frac{q(y \mid s,\theta) - q(y \mid s,0)}{\theta} - q_\theta(y \mid s,0).$$

The total variation of $\lambda(\cdot \mid s,\theta)$ is given by the following measure (Halmos (1950), pp.123)

$$\mid \lambda \mid (C \mid s,\theta) = \int_C \mid f(y \mid s,\theta) \mid \nu(dy).$$

Notice that $\mid f(y \mid s,\theta) \mid$ vanishes as θ goes to zero for each (y, s); since the convergence is dominated by $2z(y, s)$, by the LDCT $\mid \lambda \mid (C \mid s,\theta)$ converges to zero for every $C \in \mathcal{B}_Y$. In particular,

$$\lim_{\theta\to 0+} \mid \lambda \mid (Y \mid s,\theta) = \lim_{\theta\to 0+} \|\frac{Q(\cdot \mid s,\theta) - Q(\cdot \mid s,0)}{\theta} - Q_\theta(\cdot \mid s,0)\| = 0.$$

and therefore **A2** holds.

Finally, notice that

$$\begin{aligned} \|\frac{Q(\cdot \mid s,\theta) - Q(\cdot \mid s,0)}{\theta}\| &= \int_Y \mid \frac{q(y \mid s,\theta) - q(y \mid s,0)}{\theta} \mid \nu(dy) \\ &\leq \int_Y z(y,s)\nu(dy). \end{aligned}$$

By Fubini's Theorem, $M(s) = \int_Y z(y, s)\nu(dy)$ is μ-integrable, and the proof is complete. ■

Condition b) may be difficult to verify in practice, especially in problems where the densities have unbounded support. The next result can be useful in these situations.

Corollary 2 *Assume that*

a) A0 holds;

b') $q(y \mid s,\theta)$ is differentiable in θ, and either $q_\theta(y \mid s,\theta)$ is $\nu \times \mu$-integrable, or there is a $\nu\times\mu$-integrable $z(y,s)$ such that $\mid q_\theta(y \mid s,\theta) \mid \leq z(y,s)$ for every (y,s) and θ in a neighborhood of 0.

Then $V'(0+)$ exists and it is equal to zero.

Proof: A straightforward application of the Mean Value Theorem shows that b') implies b). Then the result follows from Corollary 1. ■

The following example illustrates the use of condition b'):

Example 3: Consider the information structures of the linear prediction example in Radner and Stiglitz (1984). Suppose that $Y = S = \mathbb{R}$, $q(\cdot \mid s,\theta)$ is $N(s\theta, 1-\theta^2)$, ν is the Lebesgue measure on $\mathbb{R}$, and $\mu(B) = \int_B p(s)ds$, where $p(\cdot)$ is $N(0,1)$. A little manipulation reveals that

$$q_\theta(y \mid s,\theta) = q(y \mid s,\theta)(\frac{s(y-\theta s)}{(1-\theta^2)} - \frac{(y-\theta s)^2\theta}{(1-\theta^2)^2} + \frac{\theta}{(1-\theta^2)^2}).$$

Therefore,

$$\begin{aligned}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} q_\theta(y \mid s,\theta)p(s)dyds &= \int_{-\infty}^{+\infty}(\int_{-\infty}^{+\infty} q_\theta(y \mid s,\theta)dy)p(s)ds \\ &= \int_{-\infty}^{+\infty}\frac{\theta^3}{(1-\theta^2)^2}p(s)ds \\ &= \frac{\theta^3}{(1-\theta^2)^2},\end{aligned}$$

and this is finite for $\theta \in [0,1)$. Thus, $q_\theta(y \mid s,\theta)$ is integrable, and condition b') is satisfied.

Another important case commonly found in applications is the one where signals take values on $\mathbb{R}$ and the information structure is represented by the cumulative distribution function associated with the stochastic kernel. For tractability, we focus on the case where S is any complete separable metric space but $Y = [\underline{y}, \overline{y}]$.

For each $(s,\theta) \in S \times \Theta$, let $F(\cdot \mid s,\theta) : \mathbb{R} \to [0,1]$ be the distribution function associated with $Q(\cdot \mid s,\theta)$; i.e., $F(t \mid s,\theta) = Q(\{y \leq t\} \mid s,\theta)$ for every $t \in \mathbb{R}$. The derivative of F with respect to θ will be denoted by F_θ.

Let $BV^r([\underline{y},\overline{y}])$ be the space of right-continuous functions of bounded variation $f:[\underline{y},\overline{y}]\to\mathbb{R}$ with $f(\underline{y})=0$, endowed with the total variation norm $\|f\|=V_{\underline{y}}^{\overline{y}}(f)$. Given any $f\in BV^r([\underline{y},\overline{y}])$, there exists a unique signed measure $\nu_f\in ca[\underline{y},\overline{y}]$ (the space of signed measures defined on the Borel sets of $[\underline{y},\overline{y}]$) such that $\nu_f([a,b])=f(b)-f(a)$. Actually, the spaces $BV^r([\underline{y},\overline{y}])$ and $ca[\underline{y},\overline{y}]$ are isometric, so metric relations between the elements of one space are the same as those between the corresponding elements of the other.[11]

Corollary 3 (C.D.F. Case) *Assume that*

a) A0 holds;

b) There exists a μ-integrable function $M:S\to\mathbb{R}$ such that, for every $\theta\in(0,1]$ and s,

$$\|\frac{F(\cdot\mid s,\theta)-F(\cdot\mid s,0)}{\theta}\|\leq M(s);$$

c) The following two conditions hold. For each $s\in S$ and $y\in Y$, F has a right-hand derivative at $\theta=0$; i.e.,

$$\lim_{\theta\to 0+}\frac{F(y\mid s,\theta)-F(y\mid s,0)}{\theta}=F_\theta(y\mid s,0); \tag{F1}$$

and, for each $s\in S$,

$$\lim_{\theta\to 0+}\|\frac{F(\cdot\mid s,\theta)-F(\cdot\mid s,0)}{\theta}-F_\theta(\cdot\mid s,0)\|=0. \tag{F2}$$

Then $V'(0+)$ exists and it is equal to zero.

Proof: We need to show that the conditions of Theorem 1 are satisfied. Conditions b) and F1 imply that $F_\theta\in BV^r([\underline{y},\overline{y}])$; since $BV^r([\underline{y},\overline{y}])$ and $ca[\underline{y},\overline{y}]$ are isometric spaces, **A2** and the integrability condition of Theorem 1 are obviously satisfied in the present case (they are equivalent to b) and F2). That **A1** is also satisfied can be proven as follows. Let C be a Borel set of $[\underline{y},\overline{y}]$, and let $\epsilon>0$; given any $\{\theta_n\}$ that converges to zero, we must find an N such that for all $n\geq N$, then

$$\mid\int_C d(\frac{F(y\mid s,\theta_n)-F(y\mid s,0)}{\theta_n})-\int_C dF_\theta(y\mid s,0)\mid<\epsilon.$$

[11] See Aliprantis and Border (1999), pp.364.

The left side is equal to

$$| \int I_C(y) d(\frac{F(y \mid s, \theta_n) - F(y \mid s, 0)}{\theta_n} - F_\theta(y \mid s, 0)) \mid,$$

where I_C is the indicator function of C. But this expression is less than or equal to

$$\| \frac{F(\cdot \mid s, \theta_n) - F(\cdot \mid s, 0)}{\theta_n} - F_\theta(\cdot \mid s, 0) \|$$

and this converges to zero as $n \to \infty$ by F2. Since C was an arbitrary Borel set of $[\underline{y}, \overline{y}]$, **A1** holds and the proof is complete. ■

In order to illustrate the role played by condition **A2** (F2), we present an example where all of the conditions in Theorem 1 and Corollary 3 are met except for the convergence in total variation assumption, and the nonconcavity result fails:

Example 4 (Uniform Signals): Consider the following version of the linear predictor problem: $u(a, s) = -(a - s)^2$, $S = \{s_L, s_H\}$, $q(y \mid s_L, \theta) = I_{[0,1]}$, $q(y \mid s_H, \theta) = I_{[\theta, \theta+1]}$ (see Figure 1). It is straightforward to show that **A0**, and conditions b) and F1 in Corollary 3 hold in this case; a bit of work also reveals that **A1** holds. However, condition **A2** (or F2) fails since the total variation is equal to 2 for every θ. The right derivative of the value function at $\theta = 0$ is $V'(0+) = \mu_{s_H}(1 - \mu_{s_H}) > 0$. Indeed, $V(\cdot)$ is globally *concave* in this case.

In this example, the support of the information structure is not the same for every $\theta \in \Theta$. However, this is not enough reason to make the nonconcavity disappear, as the following example reveals:

Example 5 (The Nonconcavity with Moving Supports): Let $u(a, s) = -(a - s)^2$, $S = \{s_L, s_H\}$, $q(y \mid s_L, \theta) = \max\{0, 6y(1 - y)\}$, $q(y \mid s_H, \theta) = \max\{0, 6(y - \theta)(1 - y + \theta)\}$ (see Figure 2). Tedious algebra shows that the conditions of Corollary 3 are satisfied, and $V'(0+) = 0$.

4.2 The Single Valued Case: A0 from Primitives

In this section we investigate conditions under which **A0** can be obtained from assumptions on the primitives of the decision problem.

We will find it convenient to work with the extensive form representation of the problem; i.e., given an information structure $Q(\cdot \mid \cdot, \theta)$ and prior beliefs μ, then, after observing y, the agent solves

$$\max_{a \in A} \int_S u(a, s) P(ds \mid y, \theta). \tag{8}$$

where $P(\cdot \mid y, \theta)$ is a version of the posterior beliefs of the decision maker.

In order to prove the result, it will be assumed that S is a compact metric space. Let $\mathcal{P}(S)$ be the set of probability measures on $(S, \mathcal{B}_S)$ endowed with the topology of weak convergence, and assume that $P(\cdot \mid y, \theta)$ is continuous in θ at zero for each $y \in Y$ (i.e., given any sequence $\theta_n \to 0$, then $P(\cdot \mid y, \theta_n) \overset{w}{\to} P(\cdot \mid y, 0)$). Finally, it will be assumed that A is compact and convex and $u(\cdot, s) : A \to \mathbb{R}$ is strictly concave for each $s \in S$.

The following proposition defines a class of decision problems in which **A0** is satisfied:

Proposition 3 *Let A be a compact and convex metric space, S a compact metric space, and $u(\cdot, s)$ a strictly concave function on A for each $s \in S$. If there is a version of the posterior kernel such that $P(\cdot \mid y, \theta_n) \overset{w}{\to} \mu(\cdot)$ for each y and for any sequence $\{\theta_n\}$ that converges to zero, then there is a unique decision function $d^*(y, \theta)$ that solves (8); this function is $(\mathcal{B}_Y, \mathcal{B}_A)$-measurable and continuous in θ; moreover, $d^*(y, 0) = a_0^*$ for every y. Hence, A0 is satisfied.*

Proof: For notational simplicity, let $g(a, \theta, y) = \int_S u(a, s) P(ds \mid y, \theta)$. Since $u(\cdot, s)$ is strictly concave in a for each s, $g(\cdot, \theta, y)$ is strictly concave in a for each (θ, y). Moreover, $g(\cdot, \cdot, y)$ is continuous on $A \times \{0\}$: for given any sequence $(a_n, \theta_n) \to (\overline{a}, 0)$, we have

$$\begin{aligned}
\mid g(a_n, \theta_n, y) \ &- \ g(\overline{a}, 0, y) \mid \\
&= \ \mid \int_S u(a_n, s) P(ds \mid y, \theta_n) - \int_S u(\overline{a}, s) P(ds \mid y, 0) \mid \\
&\leq \ \mid \int_S u(a_n, s) P(ds \mid y, \theta_n) - \int_S u(\overline{a}, s) P(ds \mid y, \theta_n) \mid \\
&+ \ \mid \int_S u(\overline{a}, s) P(ds \mid y, \theta_n) - \int_S u(\overline{a}, s) P(ds \mid y, 0) \mid \\
&\leq \ \int_S \mid u(a_n, s) - u(\overline{a}, s) \mid P(ds \mid y, \theta_n) \\
&+ \ \mid \int_S u(\overline{a}, s) P(ds \mid y, \theta_n) - \int_S u(\overline{a}, s) P(ds \mid y, 0) \mid . \qquad (9)
\end{aligned}$$

Since S and A are compact metric spaces and $u : A \times S \to R$ is continuous, $u(a_n, s) \to u(\overline{a}, s)$ uniformly (Dixmier (1984), Theorem 6.1.13) and therefore, given any $\epsilon > 0$, there is always an N_ϵ sufficiently large such that for all $n \geq N_\epsilon$

$$\int_S \mid u(a_n, s) - u(\overline{a}, s) \mid P(ds \mid y, \theta_n) < \epsilon \int_S P(ds \mid y, \theta_n) = \epsilon.$$

Hence, the first term in the last inequality in (9) converges to zero, while the second vanishes by weak convergence. This shows that $g(\cdot, \cdot, y)$ is continuous at every point in $A \times \ \{0\}$. That $g(a, \theta, \cdot)$ is $\mathcal{B}_Y$-measurable for each (a, θ) follows directly from the continuity of $u : A \times S \to R$ and the measurability of $P(B \mid \cdot, \theta)$ for each $B \in \mathcal{B}_S$.

Now, since A is convex and compact and $g(\cdot, \theta)$ is strictly concave, we have that for each y there is a unique solution $d^*(y, \theta)$ to problem (8). To prove continuity at $\theta = 0$, suppose to the contrary that, for some $y \in Y$, $d^*(y, \theta)$ is not continuous at $\theta = 0$. Let ρ denote the metric on the space A. Then, for some sequence $\{\theta_n\}$ tending to 0, there is an $\varepsilon > 0$ such that $\rho(d^*(y, \theta_n), d^*(y, 0)) > \varepsilon$ for all n. Since A is compact, there is a subsequence of $\{d^*(y, \theta_{n_k})\}$ with limit $\overline{d} \in A$ and $\overline{d} \neq d^*(y, 0)$. Fix $a \in A$. For every k, we must of course have $g(d^*(y, \theta_{n_k}), \theta_{n_k}) \geq g(a, \theta_{n_k})$. By the continuity of g on $A \times \{0\}$, we have $g(\overline{d}, 0) \geq g(a, 0)$. Since $a \in A$ was arbitrary, $\overline{d}$ must solve (8) for $\theta = 0$, contradicting uniqueness.

The $(\mathcal{B}_Y, \mathcal{B}_A)$-measurability of $d^*(y, \theta)$ follows from the fact that, for each θ, the conditions of the Measurable Maximum Theorem (Aliprantis and Border (1999), Theorem 17.18) are satisfied.

Finally, since $P(\cdot \mid y, \theta_n) \xrightarrow{w} \mu(\cdot)$ for any sequence $\theta_n \to 0$, we have that

$$d^*(y, \theta) = \operatorname*{argmax}_{a \in A} \int_S u(a, s) P(ds \mid y, \theta),$$

converges to

$$d^*(y, 0) = \operatorname*{argmax}_{a \in A} \int_S u(a, s) \mu(ds),$$

and the last expression is independent of y. This completes the proof of the proposition. ■

The proposition provides a set of conditions under which 'continuity and flatness' of the optimal policy at $\theta = 0$ are satisfied. Among them, weak convergence of the posterior to the prior as θ goes to zero plays a prominent role. For some special cases commonly found in applications, it is straightforward to impose assumptions on the information structure and the decision maker's prior beliefs such that the weak convergence condition is satisfied.

For each $(s, \theta) \in S \times \Theta$, let $Y(s, \theta) = \{y : q(y \mid s, \theta) > 0\} \in \mathcal{B}_Y$ be the support of $q(\cdot \mid s, \theta)$; obviously, at $\theta = 0$ we have that $Y(s, 0) = Y(s', 0) = Y_0 \subseteq Y$ for all $s, s' \in S$.

Corollary 4 *Let A be a compact and convex metric space, S a compact metric space, and $u(\cdot, s)$ a strictly concave function on A for each $s \in S$. Suppose that $Q(\cdot \mid s, \theta)$ has a density $q(\cdot \mid s, \theta)$ with respect to a σ-finite measure $\nu : \mathcal{B}_Y \to [0, \infty]$ for every $(s, \theta) \in S \times \Theta$, such that*

(i) $q(y \mid \cdot, \theta)$ is $\mathcal{B}_S$-measurable and bounded for every $(y, \theta) \in Y \times \Theta$;

(ii) For ν-almost every $y \in Y_0$, there exists a θ_y such that, for all $\theta < \theta_y$ and for every $s \in S$, $y \in Y(s, \theta)$; and

(iii) $q(y \mid s, \cdot)$ is continuous in θ at $\theta = 0$ for every $(y, s) \in Y \times S$.

Then Proposition 3 holds.

Proof: We only need to show that the weak convergence condition is satisfied. For any $y \in Y_0$ that satisfies (ii) and given any set $B \in \mathcal{B}_S$, the posterior kernel after observing $y \in Y$ is, for $\theta < \theta_y$, given by

$$P(B \mid y, \theta) = \frac{\int_B q(y \mid s, \theta) \mu(ds)}{\int_S q(y \mid s, \theta) \mu(ds)}.$$

where the integral is well-defined by (i). For the ν-measure zero set that violates (ii) and for every $y \in Y - Y_0$, set $P(B \mid y, \theta) = \mu(B)$ for all $\theta \in \Theta$. This defines a version of the posterior for every $y \in Y_0$.

Take any $\theta_n \to 0$: (i)-(iii) yield (for ν-almost every $y \in Y_0$)

$$P(B \mid y, \theta_n) = \frac{\int_B q(y \mid s, \theta_n)\mu(ds)}{\int_S q(y \mid s, \theta_n)\mu(ds)} \to \frac{\int_B q(y \mid s, 0)\mu(ds)}{\int_S q(y \mid s, 0)\mu(ds)} = \mu(B),$$

where the application of the LDCT is justified by (i), and the last equality follows from the fact that $q(y \mid, s, 0)$ is independent of s. Since the posterior converges to the prior for each Borel set B when θ goes to zero, it also converges weakly. This completes the proof. ■

Notice that this result includes as special case the 'common support assumption' which is often used in applications; i.e., $Y = \{y : q(y \mid s, \theta) > 0\}$ for every $(s, \theta) \in S \times \Theta$. It also includes the 'moving support' example presented in Example 5.

Using this corollary, it is easy to construct examples where the assumptions underlying Proposition 3 hold except for the weak convergence condition, and the value of a small amount of information is *positive*.

Example 6 (Failure of Weak Convergence): Let $u(a, s) = -(a-s)^2$, $S = \{s_L, s_H\}$, $Y = \{y_1, y_2\}$, $q(y_1 \mid s_L, \theta) = 1$, $q(y_2 \mid s_H, \theta) = g(\theta)$, $0 \leq g(\theta) \leq 1$ for every $\theta \in [0, 1]$, $g(0) = 0$, $g(1) = 1$, $g'(\theta) > 0$, and $0 < \mu_{s_L} < 1$. It is easy to show that the right derivative of the value function at $\theta = 0$ is $V'(0+) = \mu_{s_L}^2(1 - \mu_{s_L})g'(0)(s_L - s_H)^2 > 0$.

In this example, $d^*(y_2, \theta) = s_H$ for every $\theta > 0$, so a small amount of information reveals the true state with certainty when y_2 is observed. Although it is possible to find a continuous selection (in θ) from the correspondence of maximizers, one cannot find one that will also be 'flat' in y at $\theta = 0$. In terms of Proposition 3, notice that the posterior belief that the state is s_H after observing y_2 is equal to one for every $\theta > 0$, so weak convergence to the prior as θ goes to zero fails in this case; a small amount of information starting from $\theta = 0$ has a substantial effect on beliefs.

We next present an example where all the assumptions of Proposition 3 are met except for the strict concavity of $u(\cdot, s)$, and the nonconcavity result fails.

Example 7 (Failure of Single-Valued Choice): Let $u(a,s) = as$, $A = [0,1]$, $S = \{-1,1\}$, $Y = \{y_1, y_2\}$, $\mu(1) = \frac{1}{2}$, and the information structure is given by $q(y_1 \mid -1, \theta) = \frac{1}{2}$ and $q(y_1 \mid 1, \theta) = \frac{1}{2} - \theta$. In this case, it is easy to show that $V(\theta) = \theta$ and thus $V'(0+) = 1$.

Again, this is a case where **A0** fails; although any action is optimal when $\theta = 0$, the optimal decision for any $\theta > 0$ is $d^*(y_1, \theta) = 0$ and $d^*(y_2, \theta) = 1$, which reveals that 'continuity and flatness' are incompatible in this case.

As a final example illustrating Proposition 3, we consder a *noiseless* information structure: all the state-contingent signal distributions are degenerate. Noiseless structures are often used to study optimal learning.[12]

Example 8 (Noiseless Information): Let $S = [0,1], Y = \{0,1\}$, and let $Q(\{1\} \mid s, \theta) = 1$ if and only if $s > \theta$ (see Figure 3). One may verify that this information structure satisfies **A1**, **A2** and our integrability condition b) in Theorem 1. If, however, the prior has full support on $[0,1]$, then no version of the posterior converges weakly to the prior for $y = 0$. Note that $P(B \mid y, \theta) = \frac{\mu(B \cap [0,\theta])}{\mu([0,\theta])}$. Let $B \in \mathcal{B}_{\mathcal{S}}$ and $\varepsilon > 0$ satisfy $\mu(B) > 0$ and $B \cap [0, \varepsilon) = \emptyset$. Then $P(B \mid 0, \theta) = 0$ for all $\theta < \varepsilon$, so that $P(B \mid 0, \theta)$ does not converge to $\mu(B)$, and the conditions of Proposition 3 fail. We now illustrate with two utility functions that the nonconcavity may fail or hold with this family of information structures. In each example the prior belief $\mu(\cdot)$ is the Lesbesgue measure on $[0,1]$. First consider the predictor problem of Example 4, with $u(a,s) = -(a-s)^2$ and $A = [0,1]$. In this case one may verify that $V'(0) = \frac{1}{4}$, so that the marginal value of a little information is positive.[13] Second, consider the utility function $u(a,s) = a$ if $s \geq a$ and $u(a,s) = 0$ if $s < a$. This utility function describes a firm who sets a price not knowing the valuation s of a single consumer who buys at most one unit of the good (as in Aghion et al. (1991), Section 6). In this case $V(\theta) = \frac{\theta^2}{4} + \max\{\frac{1}{2}, \theta\}(1 - \max\{\frac{1}{2}, \theta\})$. Thus, $V'(0) = 0$ but $V'(1) = -\frac{1}{2}$. That is, the marginal value of a little information is zero at $\theta = 0$; however, the marginal value of lowering θ from the null information structure at $\theta = 1$ is positive.

[12] See e.g. Aghion, et. al. (1991) and the references contained therein.
[13] One may verify nonetheless that $V(\cdot)$ is convex over an interval containing 0.

5 Remarks on the Demand for Information

As Radner and Stiglitz (1984) emphasized, the nonconcavity result has important effects on the demand for information; for example, it will not be a continuous function of its price, and first order conditions need not pin down the circumstances under which information demand arises.

Since there are well-known papers in the literature that study the demand for information in static and dynamic contexts and do *not* suffer from these complications, it is important to understand why this is the case.

Kihlstrom (1974) develops a static theory of the demand for information about product quality.[14] He analyzes a consumer's problem in which the quality of one of the goods she consumes is unknown, and she can purchase different 'amounts' of information at a constant marginal cost before making her consumption decisions. In his model the demand for information is well-behaved: the marginal value of a small amount of information is positive, the quantity of information demanded is a continuous function of its price, and it is straightforward to characterize with the first order conditions of the problem the parameter values under which information demand arises. Two other papers on information demand have found the value of information to be globally concave: Freixas and Kihlstrom (1984) analyzed a specific model of demand for information about the quality of medical care; and Arrow (1985) examined the demand for information in the linear predictor model of Example 4 ($u(a,s) = -(a-s)^2$). All of these papers assumed that the decision maker has a normal prior and observes a signal that is normally distributed with mean s and variance $\frac{1}{\theta}$, i.e. $N(s, \frac{1}{\theta})$, where $\theta > 0$.[15] The rationale for this information structure is based on the familar facts that the sample mean of n i.i.d. normal random variables with mean s and unit variance is $N(s, \frac{1}{n})$; and that in this case the sample mean is a sufficient statistic for s. Thus the parameter θ represents the continuum analog of the sample size for conditionally independent normal signals with mean equal to the true state s. That the value of information can be concave for this way of indexing information structures does not contradict our theorem: the state-contingent cumulative distribution functions for the signals are not differentiable at $\theta = 0$, so that our differentiability assumption **A1** (or F1 in

[14]See also Kihlstrom (1973), and Freixas and Kihlstrom (1984).

[15]For $\theta = 0$, any null information structure can be choosen. Kihlstrom (1974) shows that the value of information in his problem converges to the value under null information as θ tends to zero.

Corollary 3) fails.

An interesting recent paper by Moscarini and Smith (1999) presents a dynamic theory of information demand, in which an individual can sample costly information about the state of the world in continuous time before stopping and taking an action. The information demand they derive is well-behaved, and the nonconcavity problem does not arise.

To be sure, their model differs in more than one way from the class of decision problems we consider, but it is instructive to explore whether the absence of the nonconcavity issue might be due to intertemporal considerations or to the information structures they considered.

Barring notational differences, Moscarini and Smith (1999) assume that the agent controls the instantaneous variance of an observation process given by the following stochastic differential equation:

$$dy_t = sdt + \frac{\sigma}{\sqrt{\theta_t}} dW_t$$

where $\{W_t\}$ is a standard Brownian motion. At each instant before the stopping time, the agent chooses the 'amount of information' θ_t that she wants to purchase.

For large intervals of time, the diffussion $\{y_t\}$ is not a Gaussian process; however, if we consider a small interval Δ, then y_t is approximately Gaussian, with $y_t \sim N(s\Delta, \frac{\sigma^2\Delta}{\theta})$.[16] This reveals that, just as with Kihlstrom (1974), the static version of the model violates **A1**, suggesting that the nonconcavity need not arise in its dynamic version either.[17]

The preceding papers show that we can sometimes avoid the nonconcavity by using the 'number of observations' to measure the amount of information. This case therefore calls for closer analysis. Observing a random variable that is normally distributed with mean s and variance $\frac{1}{\theta}$, is of course equivalent to observing a signal y given by

$$y = s\sqrt{\theta} + \varepsilon, \tag{10}$$

[16] Klebaner (1998) pp.119.

[17] Moscarini and Smith (1999) also considered an alternative observation process that yields the same results:

$$dy_t = s\theta_t dt + \sigma\sqrt{\theta_t} dW_t.$$

In this case, for a small Δ, $y_t \sim N(s\Delta\theta, \sigma^2\Delta\theta)$, which also violates **A1**.

where ε is $N(0,1)$. It is intuitively clear why the nonconcavity might fail here: the Inada condition on the conditional density of the signal y at $\theta = 0$ implies that an increase in θ from 0 spreads the signal distributions for different states apart very quickly; hence a small increase in θ from 0 is very informative. Contrast the information structure given in (10) to the oft-analyzed linear regression model,

$$y = s\theta + \varepsilon. \tag{11}$$

The optimal experimentation literature has used this model to study a firm learning about demand (e.g. Mirman, Samuelson and Urbano (1993) and Harrington (1995)), a consumer learning about product quality (e.g. Grossman, Kihlstrom and Mirman (1977)), or simply as a convenient functional form to study the long run properties of optimal learning (Kiefer and Nyarko (1989)). If $\varepsilon \sim N(0,1)$, then the conditions of Theorem 1 are met, and the nonconcavity holds (under **A0**). Thus the choice of (10) vs. (11) as the observation process has potentially dramatic consequences for analyzing the demand for information and experimentation. Unfortunately, we know precious little about how to choose functional forms for the production of information. Tentatively, however, (10) seems quite reasonable in a model of consumer learning about product quality; but it seems less plausible in a model of a firm learning about demand.

One might conjecture that the information structure given by (10) could be used to show that the value of information is concave in the number of observations for a wide class of decision problems. The next example, however, uses a simple quadratic utility function to show that the value of information can be globally *convex* in θ in this case.

Example 9 (A Convex Value of Information with Normal Sampling): Let $u(a,s) = 2a - sa^2$, $S = \{0,1\}$, $Y = \mathbb{R}$, $A = \mathbb{R}_+$ and, for $s \in \{0,1\}$, let $q(\cdot \mid s, \theta)$ be the normal density function with mean $s\sqrt{\theta}$ and unit variance. Using the extensive form of our problem, the interim value function $U : Y \times \Theta \to \mathbb{R}$ is given by $U(y,\theta) = \frac{1}{P(\{1\}|y,\theta)}$ (the optimal decision is also given by $d^*(y,\theta) = \frac{1}{P(\{1\}|y,\theta)}$). Straightforward calculations reveal that $V(\theta) = e^{\theta}$, which is globally convex.

Note that $V'(0) = 1$, so that the Radner-Stiglitz form of the nonconcavity fails here. The convexity of the value of information in θ is clearly related to the curvature of U, when viewed as a function of the posterior belief

$P(\{1\} \mid y, \theta)$. In particular, U becomes infinitely convex (in the Arrow-Pratt sense) as the posterior approaches 0. Recall that an improvement in information makes the distribution of posterior beliefs riskier. Thus each additional observation puts more weight in the tails of the distribution of posterior beliefs, including the left tail where U is becoming unboundedly convex.

It is important to notice that this example does not contradict the results of Moscarini and Smith (2000) mentioned in the introduction: they proved that the marginal value of information is eventually decreasing for a sufficiently large number of observations (which allows them to show that a well-behaved demand for information emerges for 'large quantities' or 'low prices'). They assume that both the number of actions and states are finite. Those assumptions imply in our notation that u is bounded on $A \times S$ and hence $V(\cdot)$ is bounded (even if we take its domain to be all of $\mathbb{R}_+$). A bounded, increasing function on $\mathbb{R}_+$ cannot of course be globally convex. In Example 9, u is unbounded, which permits $V(\cdot)$ to be unbounded.

6 Conclusion

We have reexamined the classic Radner and Stiglitz (1984) nonconcavity in the value of information using a general Bayesian decision framework that covers most economic applications with costly information acquisition. We have provided sufficient conditions for the existence of this nonconcavity, and they include the finite case studied by Radner and Stiglitz as a special case. We illustrated the intuition and importance of these conditions using several examples, including a discussion of some important papers on the demand for information in static and dynamic settings.

Our sufficient conditions are quite strong; yet, although they are not necessary, we have shown by examples that weakening them will not be easy. One message of the paper is thus that the Radner-Stiglitz nonconcavity emerges in a general setup only by severely constraining the set of information structures available to decision makers. Nevertheless, the smoothness and continuity conditions we use are actually weaker than those typically imposed in models of information acquisition. Moreover, as our last section suggested, even if the Radner-Stiglitz version of the nonconcavity fails, a general theorem on a globally concave value of information may yet prove elusive. As we continue to develop models of endogenous information acquisition, it seems that we will continue to confront nonconcavities in the value of information.

As a final note, we have restricted attention to single agent problems. The nonconcavity issue of course also arises in games (e.g., strategic experimentation and principal-agent models). A cursory inspection reveals that our argument directly exploits our single agent assumption (the envelope theorem explanation we give at the beginning of Section 4 is suggestive here). Hence the extension to games is not only natural but also apt to be nontrivial.

Appendix

Proof of Lemma 1: Given $\epsilon > 0$, we need to show that there exists an N_ϵ such that for every $n \geq N_\epsilon$

$$\mid \int f_n d\nu_n - \int f d\nu \mid < \epsilon.$$

By adding and subtracting $\int f_n d\nu$, the left side can be written and manipulated as follows

$$\begin{aligned} \mid \int f_n d(\nu_n - \nu) + \int (f_n - f) d\nu) \mid &\leq \mid \int f_n d(\nu_n - \nu) \mid + \mid \int (f_n - f) d\nu \mid \\ &\leq K \|\nu_n - \nu\| + \mid \int (f_n - f) d\nu \mid, \end{aligned} \tag{12}$$

where K is an upper bound of that $\mid f_n \mid$, and the inequality follows since the absolute value of the integral of a bounded measurable function with respect to a finite signed measure is less than or equal to the product of an upper bound of the integrand and the total variation of the signed measure (Royden (1988), pp.275).

Consider the first term of (12). Since ν_n converges to ν in the total variation norm, it follows that there is an N_1 such that for every $n \geq N_1$, $\|\nu_n - \nu\| < \frac{\epsilon}{2K}$.

Consider now the second term. Since $\nu = \nu^+ - \nu^-$, where ν^+ and ν^- are the positive and negative variation of ν, it follows that

$$\begin{aligned} \mid \int (f_n - f) d\nu \mid &= \mid \int (f_n - f) d\nu^+ - \int (f_n - f) d\nu^- \mid \\ &\leq \mid \int (f_n - f) d\nu^+ \mid + \mid \int (f_n - f) d\nu^- \mid . \end{aligned}$$

The integrand $f_n - f$ is bounded by $2K$ and converges to zero pointwise. Thus, by the Lebesgue Dominated Convergence Theorem there exists an N_2 such that for every $n \geq N_2$,

$$\mid \int (f_n - f) d\nu^+ \mid < \frac{\epsilon}{4}.$$

Similarly, there is an N_3 such that for every $n \geq N_3$,

$$\mid \int (f_n - f) d\nu^- \mid < \frac{\epsilon}{4}.$$

If we set $N_\epsilon = N_1 \vee N_2 \vee N_3$, then

$$| \int f_n d\nu_n - \int f d\nu |< \frac{\epsilon}{2} + \frac{\epsilon}{4} + \frac{\epsilon}{4} = \epsilon.$$

This completes the proof of the lemma. ∎

References

Aghion, P., P. Bolton, C. Harris and B. Jullien (1991) 'Optimal Learning by Experimentation', *Review of Economic Studies*, 58, 621-654.

Aliprantis, C. and K. Border (1999) *Infinite Dimensional Analysis. A Hitchhiker's guide.* Heidelberg: Springer.

Arrow, K. (1985) 'Informational Structure of the Firm', *American Economic Review Papers and Proceedings*, 75, 303-307.

Berstsekas, D. and S. Shreve (1978) *Stochastic Optimal Control: The Discrete-Time Case.* Belmont: Athena Scientific.

Billingsley, P. (1986) *Probability and Measure.* New York: John Wiley and Sons.

Bradford, D. and H. Kelejian (1977) 'The Value of Information for Crop Forecasting in a Market System: Some Theoretical Issues', *Review of Economic Studies*, 44, 519-553.

Creane, A. (1994) 'Experimentation with Heteroskedastic Noise', *Economic Theory*, 4, 275-286.

Dixmier, J. (1984) *General Topology.* New York: Springer-Verlag.

Dunford, N. and J. Schwartz (1957) *Linear Operators. Part I: General Theory.* New York: John Wiley and Sons.

Easley, D. and N. Kiefer (1988) 'Controlling a Stochastic Process with Unknown Parameters', *Econometrica*, 56, 1045-1064.

Freixas, X. and R. Kihlstrom (1984) 'Risk Aversion and Information Demand', in M. Boyer and R. Kihlstrom (eds.) *Bayesian Models of Economic Theory.* Elsevier: Amsterdam.

Grossman, S., R. Kihlstrom and L. Mirman (1977) 'A Bayesian Approach to the Production of Information and Learning by Doing', *Review of Economic Studies*, 44, 533-547.

Halmos, P. (1950) *Measure Theory.* New York: Springer.

Harrington, J. (1995) 'Experimentation and Learning in a Differentiated-Products Duopoly', *Journal of Economic Theory*, 66, 275-288.

Kiefer, N. and Y. Nyarko (1989) 'Optimal Control of an Unknown Linear Process with Learning', *International Economic Review*, 30, 571-586.

Kihlstrom, R. (1973) 'A General Theory of Demand for Information about Product Quality', *Journal of Economic Theory*, 8, 413-439.

——— **(1974)** 'A Bayesian Model of Demand for Information about Product Quality', *International Economic Review*, 15, 99-118.

———, L. Mirman and A. Postlewaite (1984) 'Experimental Consumption and the Rothschild Effect', in M. Boyer and R. Kihlstrom (eds.) *Bayesian Models of Economic Theory.* Elsevier: Amsterdam.

Klebaner, F. (1998) *Introduction to Stochastic Calculus with Applications.* London: Imperial College Press.

Laffont, J. (1985) 'On the Welfare Analysis of Rational Expectations Equilibria with Asymmetric Information', *Econometrica*, 53, 1-29.

McLennan, A. (1984) Price Dispersion and Incomplete Learning in the Long Run', *Journal of Economic Dynamics and Control*, 7, 331-347.

Mirman L., L. Samuelson and E. Schlee (1994) 'Strategic Information Manipulation in Duopolies', *Journal of Economic Theory*, 62, 363-384.

———, ——— and A. Urbano (1993) 'Monopoly Experimentation', *International Economic Review*, 34, 549- 563.

Moscarini, G. and L. Smith (1999) 'Wald Revisited: The Optimal Level of Experimentation', Mimeo, Department of Economics, University of Michigan.

———, ——— (2000) 'The Law of Large Demand for Information', Mimeo, Department of Economics, University of Michigan.

Radner, R. (1989) 'Uncertainty and General Equilibrium', in J. Eatwell, M. Milgate and P. Newman (eds.) *The New Palgrave: General Equilibrium.* New York: Norton.

———, **and J. Stiglitz (1984)** 'A Nonconcavity in the Value of Information', in M. Boyer and R. Khilstrom (eds.) *Bayesian Models of Economic Theory.* Elsevier: Amsterdam.

Royden, H. (1987) *Real Analysis.* New York: Macmillan.

Singh, N. (1985) 'Monitoring and Hierarchies: The Marginal Value of Information in a Principal-Agent Model', *Journal of Political Economy*, 93, 599-609.

Stokey, N., R. Lucas and E. Prescott (1989) *Recursive Methods in Economic Dynamics.* Cambridge: Harvard University Press.

Tonks, I. (1984) 'A Bayesian Approach to the Production of Information with a Linear Production Function', *Review of Economic Studies*, 51, 521-527.

Trefler, D. (1993) 'The Ignorant Monopolist: Optimal Learning with Endogenous Information', *International Economic Review*, 34, 565-581.

Wilson, R. (1975) 'Informational Economies of Scale', *Bell Journal of Economics*, 6, 184-195.

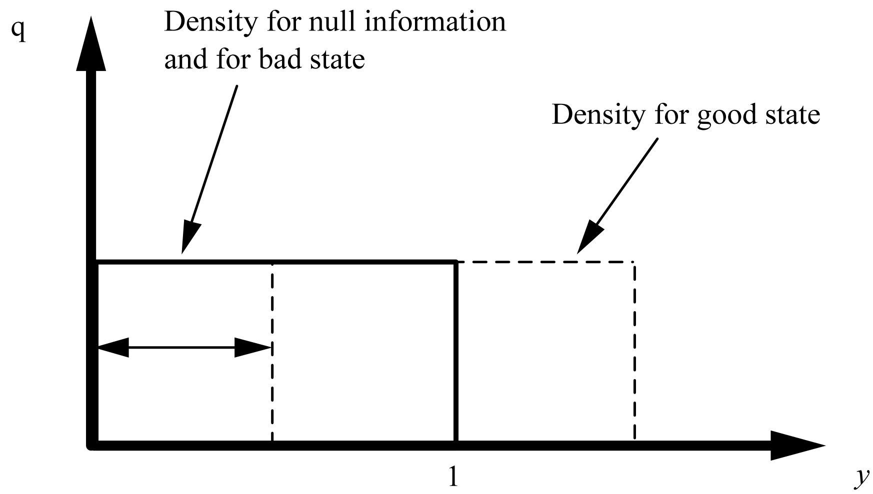

Figure 1: Example 4

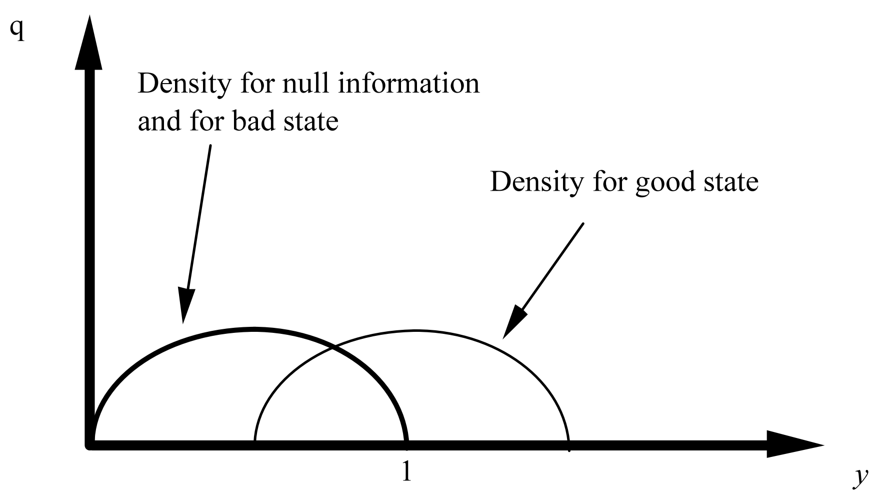

Figure 2: Example 5

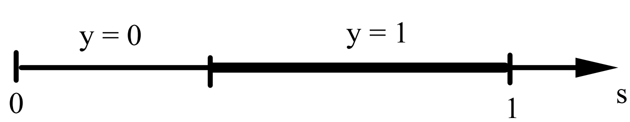

Figure 3: Example 8