Another Look at the Radner–Stiglitz Nonconcavity in the Value of Information[1]

Hector Chade and Edward E. Schlee

Department of Economics, Arizona State University, Tempe, Arizona 85287-3806
hector.chade@asu.edu, ed.schlee@asu.edu

This paper revisits the well-known result of R. Radner and J. Stiglitz (1984, *in* "Bayesian Models of Economic Theory," Elsevier, Amsterdam) 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 sufficient conditions that yield 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. *Journal of Economic Literature* Classification Number: D83.   © 2002 Elsevier Science (USA)

Key Words: value of information; information acquisition; Radner–Stiglitz nonconcavity.

1. INTRODUCTION

Is there an intrinsic nonconcavity to the value of information? A widely cited theorem of Radner and Stiglitz [27] suggests that there is. The theorem gives conditions under which 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 valueless, it must exhibit increasing marginal returns over some range. Radner and Stiglitz (henceforth RS) do present examples in which information exhibits decreasing marginal returns, so the

value of information is clearly not *always* nonconcave. Yet, the conditions of their theorem seem at first glance to be fairly innocuous smoothness and continuity assumptions. 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 corresponding to null information. They then impose two assumptions: the Markov matrix is a differentiable function of the index parameter at the zero value; and a particular selection from the correspondence of maximizers is continuous. Although these conditions may not always hold, their result is not easily dismissed as depending on exotic assumptions.

As RS 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 tend to complicate any analysis of information acquisition, and it can also have substantive consequences in applications. For example, it might preclude the existence of a competitive equilibrium (Wilson [34], Radner [26]), or the existence of a linear rational expectations equilibrium (Laffont [19]) 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 [29]).[2]

The nonconcavity has been especially vexing to the literature on active learning or experimentation.[3] In this literature, an agent takes an action in each period in the face of uncertainty about a payoff-relevant parameter. The agent observes a random signal that depends both on the action and the unknown parameter; after observing the signal, the agent updates beliefs and then chooses again. Since the signal distribution depends on the action, the agent can affect how much is learned by varying actions, sacrificing utility today to increase information available tomorrow. Thus the present action in these models acts as an index of the informativeness of the experiment that the agent observes. If the value of information is not concave in the action, then the analysis of optimal experimentation is made much more complex. Moreover, Harrington [14] and Mirman *et al.* [22] consider strategic experimentation models in which an industry of firms

---

[2] See also Arrow [2], Bradford and Kelejian [6], Lewis and Sappington [20], and Stiglitz [30] for additional remarks on the effects of the nonconcavity in the value of information.

[3] Among others, see Mirman *et al.* [23], Tonks [32], Kihlstrom *et al.* [18], Harrington [14], Easley and Kiefer [10], and Mirman *et al.* [22].

learns about demand while competing with one another. In these models, the nonconcavity means that the first period best reply mappings of firms may not be convex-valued, so that pure strategy perfect equilibria may not exist.

Besides complicating models of information acquisition, a more fundamental question remains: why should information exhibit increasing marginal returns over some range? While information as a commodity is admittedly special, it seems that there ought to exist a rich class of problems for which the value of information is concave.

All of the aforementioned applications assumed either an infinite number of signal realizations or an infinite number of states, unlike the original RS finite model. The purpose of this paper is to re-examine the RS nonconcavity—the property that a small amount of information has zero marginal value—in a general Bayesian decision problem. (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 RS nonconcavity as that which arises from a zero marginal value at null information.) Although some of our 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 clarify the role that these conditions play. Of particular interest is the assumption by RS of the existence of a selection from the correspondence of maximizers that is both continuous and constant in the signal realization at null information (our assumption **A0** below). We give 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, and illustrate the role of these conditions with several examples.

We also use our general framework to evaluate the robustness of the nonconcavity. Several important papers on information acquisition avoid the nonconcavity (e.g., Kihlstrom [16], and Moscarini and Smith [24, 25]). Most of these use (the continuum analog of) the number of conditionally independent observations from an experiment to measure the amount of information. These models suggest that the value of information might generally be a concave function of the sample size. We use a simple quadratic payoff function, however, to show that the value of information can be globally *convex* in the sample size (Example 8). Moreover, at the end of Section 5 we identify a broad class of information structure parameterizations (consistent with the Blackwell "more informative" ordering) such that the value of information must be nonconcave for some decision maker, whether or not the marginal value of a little information is zero. The conclusion we draw from our results and examples is that, although our sufficient conditions for the RS nonconcavity are strong, and one can

construct examples that yield a concave value of information, a nonconcavity in the value of information is difficult to rule out in a model of much generality.[4] Whether the reader agrees with this interpretation of our results, our hope is to stimulate thinking on the appropriate functional form restrictions to impose on information acquisition problems.

The paper is organized as follows. Section 2 sets out the general decision problem we consider. In Section 3, we state the RS theorem and provide an intuitive explanation of its assumptions as a prelude to our extension. 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. In Section 5 we use the main results to discuss some important contributions on the demand for information that do not exhibit the nonconcavity, and we also evaluate the robustness of the nonconcavity result. Section 6 concludes.

2. THE MODEL

A Bayesian agent is uncertain about the state of the world and must choose an action after observing the realization of a random variable. 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 $\mathscr{B}_S$; the measure $\mu: \mathscr{B}_S \to [0, 1]$ represents the prior beliefs of the decision maker.

• The set of signals the decision maker can observe is Y, a complete, separable metric space with Borel σ-algebra $\mathscr{B}_Y$.

• $\Theta = [0, \bar{\theta}]$ is the index set.

• For each $\theta \in \Theta$, $Q(\cdot \mid \cdot, \theta)$ is a stochastic kernel on Y given S that represents an information structure available to the agent; i.e., for each $s \in S$, $Q(\cdot \mid s, \theta): \mathscr{B}_Y \to [0, 1]$ is a probability measure, and for each $C \in \mathscr{B}_Y$, $Q(C \mid \cdot, \theta): S \to [0, 1]$ is a measurable function.[5] Different values of θ correspond to different information structures. An uninformative information structure is represented by $\theta = 0$; formally, for all $s, s' \in S$, $Q(\cdot \mid s, 0) = Q(\cdot \mid s', 0)$.

---

[4] Moscarini and Smith [25] show that if we measure the quantity of information by the number of 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. We return to this paper after presenting our main results.

[5] Throughout this paper, measurability is with respect to the appropriate Borel σ-algebra.

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

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

• $\mathscr{D}$ is the set of all measurable functions $d: Y \to A$. The set $\mathscr{D}$ contains the decision functions or strategies available to the decision maker.[6]

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

$$V(\theta) = \sup_{d \in \mathscr{D}} \int_S \int_Y u(d(y), s) \, Q(dy \,|\, s, \theta) \, \mu(ds),$$

where $V(\theta)$ is the value function of the problem, which we interpret as the value of the information structure θ. Let $D^*(\theta)$ be the correspondence of maximizers; that is,

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

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

Although we derive most of the main results using the *normal form* of the problem described above, 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.

Fix $\theta \in \Theta$; given $Q(\cdot \,|\, \cdot, \theta)$ and $\mu(\cdot)$, then one can show using standard arguments (Bertsekas and Shreve [3, Proposition 7.27]) that there exists a stochastic kernel $P(\cdot \,|\, y, \theta)$ that can be interpreted as a version of the posterior beliefs of the decision maker after observing y, and a measure $v(\cdot \,|\, \theta)$ on $\mathscr{B}_Y$ that can be interpreted as the unconditional measure for the signals when the information structure is θ; for each y, the problem is then

$$U(y, \theta) = \sup_{a \in A} \int_S u(a, s) \, P(ds \,|\, y, \theta).$$

Let $A^*(y, \theta)$ be the correspondence of maximizers. We have

$$V(\theta) = \int_Y U(y, \theta) \, v(dy \,|\, \theta). \tag{1}$$

---

[6] In the finite case, RS considered the slightly more general setup where $\mathscr{D}$ depends on θ, say $\mathscr{D}(\theta)$, and $\mathscr{D}(\theta_1) \subseteq \mathscr{D}(\theta_2)$ whenever $\theta_1 > \theta_2$. At the cost of more notation, most of our results extend to this case as well.

The extensive form representation affords a simple proof of existence of a solution and satisfaction of the measurability requirements implicit in (1).

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

    (i)   *$U(y, \theta)$ is measurable and bounded;*

    (ii) *$A^*(y, \theta)$ is nonempty and admits 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 and boundedness of $u: A \times S \to \mathbb{R}$ and measurability of the stochastic kernel imply that $g(\cdot, y, \theta): A \to \mathbb{R}$ is continuous and bounded, and $g(a, \cdot, \theta): Y \to \mathbb{R}$ is measurable. Since A is compact, (i) and (ii) follow from the Measurable Maximum Theorem (Aliprantis and Border [1, Theorem 17.18]). ∎

We assume henceforth that the maximum is attained. Until Section 4.4, we will follow RS in imposing the following assumption:

A0. There exists 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.

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: when is the marginal value of a small amount of information equal to zero? More precisely, what are conditions on the information structure which, under **A0**, imply that $V'(0+)$ exists and is zero? Note that the value function V 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. A simple, but very useful, formulation is

$$\max_{\theta \in [0, \bar{\theta}]} V(\theta) - C(\theta). \tag{2}$$

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. In addition, most standard two-period experimentation models can be written in form (2). 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 $[0, \bar{\theta}]$ say) under state s. The number θ represents a first period action that affects the distribution of the signal, and hence how much information the agent has in the second period; $V(\theta)$ then gives the maximum second period expected utility as a function of the first period action θ. Finally, the cost function equals the first period expected utility loss from choosing θ rather than the period 1 optimal choice; formally (assuming that u is the utility function in both period 1 and period 2),

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

where

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

The prototypical problem studied in the optimal experimentation literature is that of a firm learning about demand.[7]

EXAMPLE 1. A firm's demand function is $f(p, s, \varepsilon) = (\alpha - p)\, s + \varepsilon$, where p is the market price, s the state of demand, and ε the realization of a (i.i.d.) random 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. 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. A first period price of α is uninformative about s (since only "noise traders" buy at this price). Define $\theta = \alpha - p$, so that $\theta = 0$ corresponds to null information. The firm chooses θ at date 1 and observes sales of $\theta s + \varepsilon$ before choosing a date 2 price. In this case the cost function $C(\cdot)$ equals

---

[7] See, among others, McLennan [21], Mirman *et al.* [23], Treffler [33], and Creane [7].

the expected profit loss 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, if $V'(0+) = 0$ and the cost function $C(\cdot)$ is increasing with $C'(0+) > 0$, then the objective function in (2) cannot be concave if information has positive net value for some $\theta > 0$. Indeed, $V(\theta) - C(\theta)$ cannot even be quasiconcave on $[0, \bar{\theta}]$.[8] From the perspective of these applications, we can rephrase our question as determining whether the objective function in (2) can be concave for a cost function with positive marginal cost at $\theta = 0$. RS considered 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:

PROPOSITION 2. *Assume that*

   (a)   A0 *holds;*

   (b) $q(y \mid s, \theta)$ *is differentiable with respect to θ at $\theta = 0$.*

Then $D^+V(0) = \lim \sup_{\theta \to 0+} \frac{V(\theta) - V(0)}{\theta} \leqslant 0$.

In words, if $V'(0+)$ exists, it must be nonpositive. To motivate our extension of this result, consider the two assumptions more closely. Condition (b) ensures that the information structure varies smoothly with θ around 0; intuitively, it ensures that information does not increase too rapidly around null information. Regarding **A0**, one assumption that helps ensure the existence of the desired continuous selection is that the optimal choice is single-valued in the posterior (which follows if A is convex and $u(\cdot, s): A \to \mathbb{R}$ is strictly concave). Intuitively, it is easy to see how the conclusion can fail if the optimal choice is not single-valued. 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, so that even a small increase in information can have a positive marginal value.[9] As it turns out, however, strict concavity of u and condition (b) are not sufficient to yield **A0** in the finite case.

---

[8] Hence, the decision maker's *preferences* over elements of $[0, \bar{\theta}]$ (represented by $V - C$) cannot be convex. In that sense our result is an *ordinal* one, a point to which we return in Section 4.

[9] 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 is not single-valued.

Recall that the posterior belief that the state is s after observing y for information structure indexed by θ is given by: $P(\{s\} \mid y, \theta) = \frac{\mu_s q(y \mid s, \theta)}{\sum_{s \in S} \mu_s q(y \mid s, \theta)}$, where $\mu_s = \mu(\{s\})$ for all $s \in S$. Assumption (b) does not 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 $P(\{s_H\} \mid y_1, \theta)$ does not converge to the prior of $\frac{1}{2}$ as θ tends to 0.

Intuitively, without continuity of the posterior in θ at $\theta = 0$, small changes in θ yield large changes in the distribution of posterior beliefs, which may have a positive marginal value. One way to rule out this discontinuity in the finite case is to assume that the null information structure has full support on the signals for each state. (In Section 4 we extend this insight to rationalize **A0** in the general model, without requiring full support on the signals.)

In sum, the conclusion of Proposition 2 holds if u is strictly concave, and if both (b) holds and $q(y \mid s, 0) > 0$ for all $(y, s) \in Y \times S$. Intuitively, the former helps ensure that even small increases in information are not too valuable; and the latter ensures that information does not increase too quickly as θ increases from 0. We now turn to the task of extending Proposition 2 to the more general setup described in Section 2 and to identify some important specifications for which the nonconcavity fails.

4. MAIN RESULTS

Before plunging into the details of our extension, let us give an informal overview of our argument. Suppose for a moment that there is a selection $d^*(y, \theta)$ such that both $u(d^*(y, \theta), s)$ 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

$$\begin{aligned}
V'(0) &= \int_S \int_Y u(a_0^*, s) \, Q_\theta(dy \mid s, 0) \, \mu(ds) \\
&= \int_S u(a_0^*, s) \left(\int_Y Q_\theta(dy \mid s, 0) \right) \mu(ds) \\
&= 0,
\end{aligned} \tag{3}$$

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 words, the marginal value of information is zero if we start from null information (which implies that V cannot be globally concave if $V(\theta) > V(0)$ for some $\theta \in (0, \bar{\theta}]$). We derived this result, of course, 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. We now rigorously derive the conclusion of (3), eschewing the differentiability assumption on the selection.

4.1. *Generalization of the Radner–Stiglitz Theorem*

Let $ca(\mathscr{B}_Y)$ be the space of finite signed measures on $(Y, \mathscr{B}_Y)$, and endow it with the total variation norm $\|\lambda\| = |\lambda|(Y)$ (Halmos [13, pp. 122–123]). We will impose the following smoothness assumption, an extension of the differentiability condition (b) of Proposition 2.

A1.   For each $s \in S$ and $C \in \mathscr{B}_Y$,

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

exists in $\mathbb{R}$.

By Corollary 4 in Dunford and Schwartz [9, p. 160] $Q_\theta(\cdot \mid s, 0)$ is an element of $ca(\mathscr{B}_Y)$. 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** holds in the total variation norm (which follows automatically if Y is finite).

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:

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

$$\lim_{n \to \infty} \int f_n \, dv_n = \int f \, dv.$$

Proof. Appendix. ∎

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 $\theta \in (0, \bar{\theta}]$ and $s \in S$,*

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

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

Proof. We first show that $\limsup_{\theta \to 0+} \frac{V(\theta) - V(0)}{\theta} \leqslant 0$. As in RS, write $\frac{V(\theta) - V(0)}{\theta} = \frac{T_1(\theta)}{\theta} + \frac{T_2(\theta)}{\theta}$, where

$$T_1(\theta) = \int_S \int_Y u(d^*(y, \theta), s)\, Q(dy\,|\,s, \theta)\, \mu(ds)$$

$$- \int_S \int_Y u(d^*(y, \theta), s)\, Q(dy\,|\,s, 0)\, \mu(ds),$$

$$T_2(\theta) = \int_S \int_Y u(d^*(y, \theta), s)\, Q(dy\,|\,s, 0)\, \mu(ds)$$

$$- \int_S \int_Y u(a_0^*, s)\, Q(dy\,|\,s, 0)\, \mu(ds).$$

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

Consider

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

Assumption A1 ensures that this integral is well-defined for every θ.[10] Moreover, since

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

---

[10] This follows from Stokey *et al.* [31, Theorem 8.4], which also holds for signed kernels like the ones considered here (just decompose the signed kernel into its positive and negative variation).

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

We now prove that

$$\lim_{\theta \to 0+} \frac{T_1(\theta)}{\theta} = \int_S \int_Y u(a_0^*, s) \, Q_\theta(dy \mid s, 0) \, \mu(ds). \tag{4}$$

Take any sequence θ_n converging to 0, and let

$$h_n(s) = \int_Y u(d^*(y, \theta_n), s) \left(\frac{Q(dy \mid s, \theta_n) - Q(dy \mid s, 0)}{\theta_n} \right),$$

$$h(s) = \int_Y u(a_0^*, s) \, Q_\theta(dy \mid s, 0).$$

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

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

where $B < \infty$ is such that $|u(a, s)| \leqslant B$ (Royden [28, 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, (4) holds.

If we could show that $\int_Y Q_\theta(dy \mid s, 0) = 0$, it would then follow by (3) 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 (set $f_n = 1$ for every n) yields

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

Hence,

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

Finally, since $V(\theta) - V(0) \geqslant 0$, we have that $\liminf_{\theta \to 0+} \frac{V(\theta) - V(0)}{\theta} \geqslant 0$. Therefore $V'(0+)$ exists and it is equal to zero. ∎

Remark. Notice that condition (b) can be slightly relaxed and assumed to hold only for θ sufficiently close to 0; one simply needs to add this qualification in each step of the proof that uses (b) (e.g., when appealing to the LDCT). This also applies to all the results in the remainder of the paper.

Theorem 1 raises several questions, namely,

    (i)   How can one verify conditions **A1** and **A2** in applications?

    (ii)   To what extent is the nonconcavity an artifact of the units with which we measure information? Can we not undo the nonconcavity by changing the units?

    (iii)   Can **A0** be derived from assumptions imposed separately on the primitives of the model?

We next address each of these important issues in turn.

4.2. *Verification of Conditions A1 and A2*

Checking that assumptions **A1** and **A2** hold could 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 $v: \mathscr{B}_Y \to [0, \infty]$ 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 v that is also μ-integrable. That is, for every $C \in \mathscr{B}_Y$ and $(s, \theta) \in S \times \Theta$,

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

In particular, this includes the important case in applications where Y is a (Borel) subset of $\mathbb{R}^n$ (endowed with the Euclidean metric), v 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 v is just $Q(\cdot \mid s, 0)$. Finally, it includes the countable case if v 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 v for each (s, θ). Obviously, this subsumes the finite case considered by RS.

The value of information structure θ in this case becomes

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

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 $v \times \mu$-integrable function $z(y, s)$ such that $|\frac{q(y \mid s, \theta) - q(y \mid s, 0)}{\theta}| \leqslant z(y, s)$ for every (y, s) and $\theta > 0$.*

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

Proof.   Appendix.   ∎

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 $v \times \mu$-integrable, or there is a $v \times \mu$-integrable $z(y, s)$ such that $|q_\theta(y \mid s, \theta)| \leqslant z(y, s)$ for every (y, s) and θ.*

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 usefulness of condition (b'), and also that the dominating function could depend on θ:

EXAMPLE 3.   Consider the information structures of the linear prediction example in RS. Suppose that $Y = S = \mathbb{R}$, $q(\cdot \mid s, \theta)$ is $N(s\theta, 1 - \theta^2)$,

$\theta \in [0, 1]$, v 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

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

and

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

Also,

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} z(y, s, \theta) \, p(s) \, dy \, ds = \frac{2}{\pi} \frac{1}{(1 - \theta^2)^{\frac{1}{2}}} + \frac{2\theta}{(1 - \theta^2)} < \infty,$$

for θ close to zero, and

$$\lim_{\theta \to 0} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} z(y, s, \theta) \, p(s) \, dy \, ds = \frac{2}{\pi} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} z(y, s, 0) \, p(s) \, dy \, ds.$$

Using Theorem 17 in Royden [28, p. 91], it can be shown that q_θ is integrable and thus 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 (c.d.f.) associated with the stochastic kernel.[11] For tractability, we focus on the case where S is any complete separable metric space but $Y = [\underline{y}, \bar{y}]$.

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

Let $BV'([\underline{y}, \bar{y}])$ be the space of right-continuous functions of bounded variation $f \colon [\underline{y}, \bar{y}] \to \mathbb{R}$ with $f(\underline{y}) = 0$, endowed with the total variation norm $\|f\| = V_{\underline{y}}^{\bar{y}}(f)$. Given any $f \in BV'([\underline{y}, \bar{y}])$, there exists a unique signed measure $v_f \in ca[\underline{y}, \bar{y}]$ (the space of signed measures defined on the Borel sets of $[\underline{y}, \bar{y}]$) such that $v_f([a, b]) = f(b) - f(a)$; moreover, the spaces $BV'([\underline{y}, \bar{y}])$ and $ca[\underline{y}, \bar{y}]$ are isometric.[12]

---

[11] This case is useful for two reasons. First, the stochastic kernel may not be absolutely continuous with respect to any σ-finite measure v, even if the c.d.f. is differentiable. Second, even if such a σ-finite measure exists, the conditions of Corollary 1 may be hard to verify if that measure is not Lebesgue.

[12] See Aliprantis and Border [1, p. 364].

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 $\theta \in (0, \bar{\theta}]$ and s,*

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

(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.,*

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

and, for each $s \in S$,

$$\lim_{\theta' \to 0+} \left\| \frac{F(\cdot \mid s, \theta') - F(\cdot \mid s, 0)}{\theta'} - F_{\theta}(\cdot \mid s, 0) \right\| = 0. \qquad \text{(F2)}$$

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

Proof.   Appendix.   ∎

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 Fig. 1), and $0 < \mu_{s_H} < 1$. 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 $V'(0+) = (s_H - s_L)^2 \mu_{s_H}(1 - \mu_{s_H}) > 0$. Indeed, $V(\cdot)$ is globally *concave* in this case.

In the example, the signal distributions have different support for each state, the posterior belief has precisely three point support (0, 1, or the prior) for any $0 < \theta < \bar{\theta}$, and the probability that the state is learned is a linear function of θ. Intuitively, small increases in θ from 0 provide lots of information. **A2**—or something like it—must be imposed to rule out this example. The next example shows that **A2** and the nonconcavity are consistent with signal distributions with moving supports.
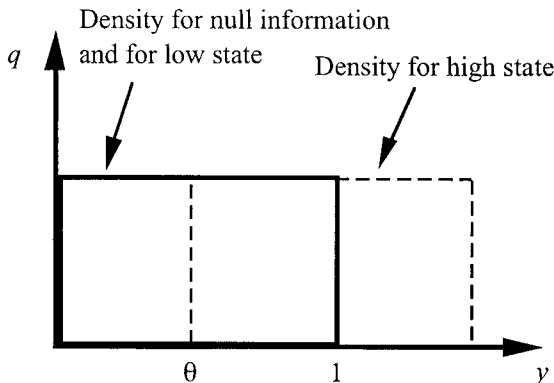
FIG. 1. Example 4 (Uniform densities).

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 Fig. 2), and $0 < \mu_{s_H} < 1$. Tedious algebra shows that the conditions of Corollary 3 are satisfied, and $V'(0+) = 0$.

4.3. Changes in the Units of Information

One concern the reader might have about the nonconcavity is that it is simply a matter of the units with which we measure the quantity of information: can't we dispose of the nonconcavity by redefining the units? Theorem 1, after all, simply deals with the gross value of information, V, ignoring any costs. If, however, we embed V in a problem of costly information acquisition (as in (2)), then the economically important question is whether the objective function $V - C$ is quasiconcave (i.e., whether preferences over elements of $[0, \bar{\theta}]$ are convex). The property that $V'(0+) = 0$
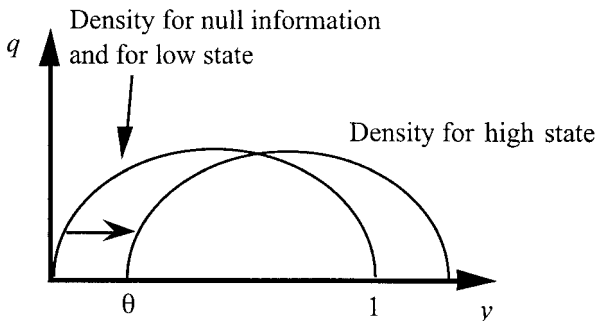

FIG. 2. Example 5 (Moving supports).

is simply a convenient way to prove that $V - C$ cannot be quasiconcave if $C'(0+) > 0$ and information has positive net value for some θ. The following result helps clarify that the failure of quasiconcavity is immune to (monotone) transformations of the units of information.

PROPOSITION 3. *Let* $\psi: [0, \bar{\theta}] \to R$, *let* $T: [0, \bar{\theta}] \to [0, \bar{\theta}]$ *be strictly increasing and onto and define* $\hat{\psi}: [0, \bar{\theta}] \to R$ *by* $\hat{\psi}(\hat{x}) = \psi(T^{-1}(\hat{x}))$ *for all* $\hat{x} \in [0, \bar{\theta}]$. *If* $\hat{\psi}$ *is quasiconcave, then* ψ *is quasiconcave.*

Proof. Suppose that ψ is not quasiconcave. Then there is a $k \in \mathbb{R}$ and numbers x_1, x_2, λ in $[0, 1]$ such that $\psi(x_i) \geqslant k$ for $i = 1, 2$ but $\psi(\lambda x_1 + (1 - \lambda) x_2) < k$. Define $\hat{x}_i = T(x_i)$ for $i = 1, 2$. Then $\hat{\psi}(\hat{x}_i) = \psi(T^{-1}(T(x_i))) = \psi(x_i) \geqslant k$ for $i = 1, 2$. Define $\mu \in [0, 1]$ by

$$\mu = \frac{T(\lambda x_1 + (1 - \lambda) x_2) - T(x_2)}{T(x_1) - T(x_2)}.$$

which is possible since $x_1 \neq x_2$, T is strictly increasing, and $\lambda \in (0, 1)$. We have

$$
\begin{aligned}
\hat{\psi}(\mu\hat{x}_1 + (1 - \mu) \hat{x}_2) &= \hat{\psi}(\mu T(x_1) + (1 - \mu) T(x_2)) \\
&= \hat{\psi}(T(\lambda x_1 + (1 - \lambda) x_2)) \\
&= \psi(T^{-1}(T(\lambda x_1 + (1 - \lambda) x_2))) \\
&= \psi(\lambda x_1 + (1 - \lambda) x_2).
\end{aligned}
$$

Hence, $\hat{\psi}(\mu\hat{x}_1 + (1 - \mu)\hat{x}_2) = \psi(\lambda x_1 + (1 - \lambda) x_2) < k$, so that $\hat{\psi}$ is not quasiconcave. ∎

COROLLARY 4. *If* $\theta \mapsto V(\theta) - C(\theta)$ *is not quasiconcave on* $[0, \bar{\theta}]$, *then* $\hat{\theta} \mapsto V(T^{-1}(\hat{\theta})) - C(T^{-1}(\hat{\theta}))$ *will not be quasiconcave on* $[0, \bar{\theta}]$ *for any strictly increasing, onto function* $T: [0, \bar{\theta}] \to [0, \bar{\theta}]$.

An implication of the corollary is that if $V'(0+) = 0$, $C'(0+) > 0$, and $V(\theta) - C(\theta) > V(0) - C(0)$ for some $\theta \in (0, \bar{\theta}]$, then no (monotone) redefining of the units of information can do away with the non-quasiconcavity. In that sense, our result is an ordinal property, not merely an artifact of the units in which information is measured.

4.4. *The Single Valued Case:* A0 *from Primitives*

We now give conditions on the primitives of the problem that imply condition **A0**. 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

$$\max_{a \in A} \int_S u(a, s) \, P(ds \mid y, \theta), \tag{5}$$

where $P(\cdot \mid y, \theta)$ is a version of the posterior beliefs of the decision maker. To prove the result, we will assume that S is a compact metric space. Let $\mathscr{P}(S)$ be the set of probability measures on $(S, \mathscr{B}_S)$ endowed with the topology of weak convergence.

PROPOSITION 4. *Let A be a compact and convex metric space, S a compact metric space, Y a complete and separable 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 solution $d^*(y, \theta)$ to (5); the function $d^*(\cdot, \theta)$ is measurable and $d^*(y, \cdot)$ is continuous at $\theta = 0$; moreover, $d^*(y, 0) = a_0^*$ for every y. Hence,* A0 *is satisfied.*[13]

Proof. Appendix. ∎

The proposition provides 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 such that the weak convergence condition is satisfied.

For each $(s, \theta) \in S \times \Theta$, let $Y(s, \theta) = \{y : q(y \mid s, \theta) > 0\} \in \mathscr{B}_Y$ 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 5. *Let A, S, Y and u satisfy the conditions of Proposition 4. Suppose that $Q(\cdot \mid s, \theta)$ has a density $q(\cdot \mid s, \theta)$ with respect to a σ-finite measure $v: \mathscr{B}_Y \to [0, \infty]$ for every $(s, \theta) \in S \times \Theta$, such that*

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

(ii)   *For v-almost every $y \in Y$, there exists a $\theta_y \in (0, \bar{\theta}]$ such that either (a) for all $\theta < \theta_y$ and for every $s \in S$, $y \in Y(s, \theta) \cap Y_0$; or (b) for all $\theta < \theta_y$ and for every $s \in S$, $y \notin Y(s, \theta) \cap Y_0$.*

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

Then Proposition 4 holds.

---

[13] Strict concavity of $u(\cdot, s)$ (on the convex space A) ensures that (5) has a unique solution; it can be replaced by any assumption that implies such uniqueness.

Proof. Appendix. ∎

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' cases in Examples 4 and 5.

Using Corollary 5, it is easy to construct examples where the assumptions underlying Proposition 4 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 $u(a, s) = -(a - s)^2$, $S = \{s_L, s_H\}$, $Y = \{y_1, y_2\}$, $q(y_1 \mid s_L, \theta) = 1$, $q(y_2 \mid s_H, \theta) = g(\theta)$, $0 \leqslant g(\theta) \leqslant 1$ 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 4, 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. In terms of Corollary 5, this example violates condition (ii).

In the next example, all the assumptions of Proposition 4 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 $q(y_1 \mid 1, \theta) = \frac{1}{2} - \theta$, with $\theta \in [0, \frac{1}{2}]$. In this case, it is easy to show that $V(\theta) = \frac{\theta}{2}$ and thus $V'(0+) = \frac{1}{2}$.

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.

If we combine Corollaries 1 and 5, then the following assumptions on the primitives define a *class of problems* where there is a nonconcavity in the value of information, a class which obviously includes a host of information acquisition models:

COROLLARY 6. *Let A, S, Y and u satisfy the conditions of Proposition* 4. *Suppose that $Q(\cdot \mid s, \theta)$ has a density $q(\cdot \mid s, \theta)$ with respect to a σ-finite*

measure $v: \mathcal{B}_Y \to [0, \infty]$ *for every* $(s, \theta) \in S \times \Theta$, *and suppose that* $q(\cdot \mid s, \theta)$ *satisfies condition (b) of Corollary 1 and (i)–(ii) of Corollary 5. Then* $V'(0+)$ *exists and it is equal to zero.*

To sum up, our extension imposes two sorts of conditions, those on the decision maker (i.e. the utility function) and those on how the available information structures are parameterized. The intuition behind these conditions is clearly analogous to that given in the finite case in Section 3: the former ensure that the optimal action is continuous in beliefs, so that small changes in beliefs do not result in large jumps in the action (hence the decision maker does not value small changes in information too much); and the latter ensure that information does not increase too much as θ rises from 0.

5. REMARKS ON THE DEMAND FOR INFORMATION

As RS pointed out, the nonconcavity 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.

Given that there are prominent papers in the literature on information demand that do *not* suffer from these complications, we should explain which assumptions of ours these models violate. Since we will be considering the behavior of information value globally (and not just in a neighborhood of null information), increases in θ in this section will always represent Blackwell improvements in the information structure (so that V is increasing in θ for *all* decision problems).

Kihlstrom [16] (see also [17]) analyzes a consumer's problem in which the quality of one of the goods is unknown, and she can purchase different amounts of information at a constant marginal cost before making her consumption decision. 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 [11] analyzed a specific model of demand for information about the quality of medical care; and Arrow [2] examined the demand for information in a linear predictor model. All of these papers assume 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 $\theta > 0$. (For $\theta = 0$, choose any null

information structure.) Recalling that the sample mean of n i.i.d. normal random variables with mean s and unit variance is distributed $N(s, \frac{1}{n})$ (and that the sample mean is a sufficient statistic for s), the parameter θ thus is 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 is concave in these models 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 Corollary 3) fails.

An interesting recent paper by Moscarini and Smith [24] 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.

Clearly, their model differs in more than one way from the class of decision problems we consider, but it is instructive to examine the information structures they assume. Barring notational differences, Moscarini and Smith [24] assume that the agent controls the instantaneous variance of an observation process given by the stochastic differential equation

$$dy(t) = s\, dt + \frac{\sigma}{\sqrt{\theta(t)}}\, dW(t),$$

where $\{W(t)\}$ is a standard Brownian motion. In each period $[t, t+dt)$ before stopping, the agent chooses the amount of information $\theta(t)$ that the agent wants to purchase. Notice that $dy(t) \sim N(sdt, \frac{\sigma^2}{\theta(t)}\, dt)$; this reveals that, just as with Kihlstrom [16], the static version of the model violates **A1**, suggesting that the nonconcavity need not arise in its dynamic version either.[14]

---

[14] Another example of a well-behaved dynamic model of information demand is the team problem in Bolton and Harris [5]. In each period $[t, t+dt)$, the social planner chooses the proportion θ_i of the current period, $i = 1, ..., I$, that each agent devotes to play a risky arm whose instantaneous mean depends on the unknown state of the world $s \in \{l, h\}$. The signal process is given by $dy_i(t) = \theta_i s\, dt + \sqrt{\theta_i}\, \sigma\, dW_i(t)$. The planner maximizes the average payoff of the players, which depends on $\sum_{i=1}^{I} \theta_i$. Any solution has each player choosing the same proportion θ; that is, $\sum_{i=1}^{I} \theta_i = I\theta$. They show that the continuation payoff from next period onward (the analogue to our $V(\theta)$; see Section 3) is linear and nondecreasing in θ (with $V'(\theta) = 0$ a possibility). Thus, the value of information is always (weakly) concave. As with the previous example, the information structures they use violate **A1**.

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 equivalent to observing a signal y given by

$$y = s\sqrt{\theta} + \varepsilon, \tag{6}$$

where ε is $N(0, 1)$. It is intuitively clear why the RS 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 (6) to the oft-analyzed[15] linear regression model,

$$y = s\theta + \varepsilon. \tag{7}$$

If $\varepsilon \sim N(0, 1)$, then the conditions of Theorem 1 are met, and the nonconcavity holds (under **A0**). Thus the choice of (6) vs (7) as the observation process can have dramatic consequences for information acquisition. Unfortunately, we know precious little about how to choose functional forms for the production of information. Tentatively, however, (6) seems quite reasonable in a model of consumer learning about product quality; but less so in a model of a firm learning about demand.

In light of these examples, one might conjecture that the information structure given by (6) 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 8 (A convex value of information with normal sampling).   Let $u(a, s) = 2a - sa^2$, $S = \{0, 1\}$, $Y = \mathbb{R}$, $A = \mathbb{R}_+$, $0 < \mu(1) < 1$ 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 $U(y, \theta) = \frac{1}{P(\{1\} \mid y, \theta)}$ (the optimal decision is also given by $d^*(y, \theta) = \frac{1}{P(\{1\} \mid y, \theta)}$). Straightforward calculations reveal that $V(\theta) = 2 - \mu(1) + \frac{(1 - \mu(1))^2}{\mu(1)} e^{\theta}$, which is globally convex.

Note that $V'(0) = \frac{(1 - \mu(1))^2}{\mu(1)} > 0$, so that the RS form of the nonconcavity fails, even though V is strictly convex.

---

[15] See Mirman *et al.* [23], Harrington [14], Grossman *et al.* [12], and Kiefer and Nyarko [15].

This example, we should stress, does not contradict the results of another paper on the nonconcavity by Moscarini and Smith [25]: they prove 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. A bounded, increasing (and nonconstant) function on $\mathbb{R}_+$ cannot be globally convex. In Example 8, u is unbounded, which permits $V(\cdot)$ to be unbounded. [16]

Example 8 shows that there is no hope for proving a theorem saying that, with normal sampling, the value of information will be concave for all decision problems and priors. We conclude this section with an argument illustrating the difficulty more generally. Suppose that $S = \{s_1, s_2\}$, that the agent's prior belief that the state is s_2 is between 0 and 1, and that $u(\cdot, s)$ is strictly concave for each $s \in S$. Posterior beliefs can thus be described by a real number p giving the probability that s_2 holds. By standard envelope arguments, $U: [0, 1] \to \mathbb{R}$ (the interim value function from the extensive form of the problem viewed as a function of the posterior) is convex and differentiable in p; moreover, for any such function U, there exists a decision problem that generates U as its value function. From an *ex ante* viewpoint, the posterior belief that $s = s_2$ is a random variable whose distribution depends on θ. Letting $G(\cdot; \theta)$ denote the cumulative distribution function of the posterior, [17] we have $V(\theta) = \int_0^1 U(p) \, dG(p; \theta)$. Integrating by parts twice yields

$$V(\theta) = U(1) - U'(1) \int_0^1 G(\omega; \theta) \, d\omega + \int_0^1 \left[\int_0^p G(\omega; \theta) \, d\omega \right] dU'(p).$$

Since the mean of the posterior always equals the prior, $\int_0^1 G(\omega; \theta) \, d\omega$ does not depend on θ. It follows that V is concave for all differentiable, convex U if and *only if* $\int_0^p G(\omega; \theta) \, d\omega$ is concave in θ for every p in $[0, 1]$: for the only if part, observe that if $\int_0^{\hat{p}} G(\omega; \theta) \, d\omega$ is not concave in θ for some $\hat{p}$ in

---

[16] Although we must use an unbounded utility function to get a strictly convex value of information, we can modify the set of states in Example 8 so that utility is bounded, yet the value of information is not concave: Let $S_n = \{1/n, 1\}$ replace S in the example, with all else unchanged, so that u is bounded on $A \times S_n$ for all n. It is straightforward to show that the corresponding value function V_n converges pointwise to $V(\theta) = 1 + (1 - \mu(1))(1 + e^\theta)/\mu(1)$; hence for n large enough, V_n is not concave in the sample size, despite the assumption of normal sampling.

[17] Formally, $G(p; \theta) = \sum_i \int_{\{y \mid P(y, \theta) \leqslant p\}} Q(dy | s_i, \theta) \, \mu_i$, where P is a version of the posterior and μ_i is the prior probability that $s = s_i$.
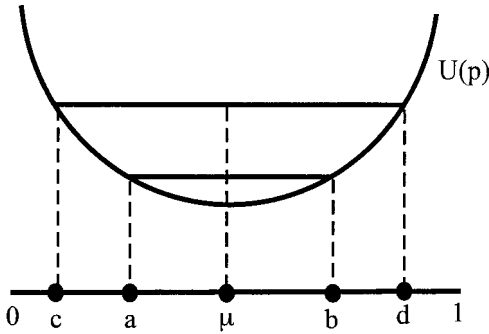
FIG. 3. A nonconcave value of information. If the posterior distribution support expands (from (a, b) to (c, d)), choose U to be sufficiently flat on (a, b) and sufficiently convex on (c, d)–(a, b).

$(0, 1)$, then it will not be concave in θ on an interval around $\hat{p}$; now simply choose U' to put most of the mass on this interval to yield V is not concave.

Under what conditions will $\int_0^p G(\omega; \theta)\, d\omega$ fail to be concave in θ for some p? One sufficient condition is that increases in θ enlarge the support of the distribution of posteriors in the following sense: for some θ', p' in $(0, 1)$, $G(p, \theta') = 0$ for all $p \leqslant p'$ but $G(p', \theta'') > 0$ for some $\theta'' > \theta'$.[18] Such information structures are hardly pathological. If, for example, the likelihood ratio, $f(y\,|\,s_H, \theta)/f(y\,|\,s_L, \theta)$ is uniformly bounded away from 0 for each θ, and the information structure tends to perfect information as θ increases, then the expanding posterior support condition holds. Moreover, for a finite number of signals, expanding support of the distribution of posteriors is clearly the typical case (see Fig. 3 for an illustration). While not definitive, the foregoing argument suggests that a nonconcavity in the value of information is difficult to avoid in a model of much generality.

6. CONCLUSION

We have reexamined the classic Radner–Stiglitz [27] nonconcavity in the value of information using a general Bayesian decision framework. Broadly, our extension imposes two kinds of restrictions: those on the decision maker (the utility is continuous in the action and state and strictly

[18] Since increases in θ represent Blackwell information improvements, the distribution of the posterior undergoes a mean-preserving increase in risk as θ increases. Letting $\Psi(p, \theta) = \int_0^p G(\omega, \theta)\, d\omega$, we must have $0 = \Psi(p', 0) = \Psi(p', \theta') < \Psi(p', \theta'')$, so that $\Psi(p', \cdot)$ is not concave.

concave in the action) and those on how the available information structures are parameterized (**A1, A2** and the weak convergence condition of Proposition 4). Intuitively, the former ensures that the optimal action is unique, hence continuous, in beliefs, so that small changes in beliefs result in only small changes in the action (and hence the decision maker does not value small increases in information too much); and the latter ensure that information does not increase too rapidly in the information parameter around null information.

Our sufficient conditions are 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 only by severely constraining the set of information structures available to decision makers. Nevertheless, our conditions are weaker than those often imposed in models of information acquisition. Moreover, as our last section suggests, even if the Radner–Stiglitz version of the nonconcavity fails, a general theorem on a globally concave value of information is likely to 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, and hence the complications reviewed in the Introduction.

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 $n \geqslant N_\epsilon$

$$\left| \int f_n \, dv_n - \int f \, dv \right| < \epsilon.$$

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

$$\left| \int f_n d(v_n - v) + \int (f_n - f) \, dv \right| \leqslant \left| \int f_n d(v_n - v) \right| + \left| \int (f_n - f) \, dv \right|$$

$$\leqslant K \, \|v_n - v\| + \left| \int (f_n - f) \, dv \right|, \qquad (8)$$

where K is an upper bound of the sequence $|f_n|$, 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 [28, p. 275]).

Consider the first term of (8). Since v_n converges to v in the total variation norm, it follows that there is an N_1 such that for every $n \geqslant N_1$, $\|v_n - v\| < \frac{\epsilon}{2K}$.

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

$$\left| \int (f_n - f) \, dv \right| = \left| \int (f_n - f) \, dv^+ - \int (f_n - f) \, dv^- \right|$$

$$\leqslant \left| \int (f_n - f) \, dv^+ \right| + \left| \int (f_n - f) \, dv^- \right|.$$

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 $n \geqslant N_2$,

$$\left| \int (f_n - f) \, dv^+ \right| < \frac{\epsilon}{4}.$$

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

$$\left| \int (f_n - f) \, dv^- \right| < \frac{\epsilon}{4}.$$

If we set $N_\epsilon = N_1 \vee N_2 \vee N_3$, then $|\int f_n dv_n - \int f \, dv| < \frac{\epsilon}{2} + \frac{\epsilon}{4} + \frac{\epsilon}{4} = \epsilon$.

This completes the proof of the lemma. ∎

Proof of Corollary 1.   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 $\theta > 0$,

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

Since the integrand is dominated, the LDCT implies

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

Define

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

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

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

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

for every $C \in \mathscr{B}_Y$, 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 [13, p. 123])

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

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

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

and therefore **A2** holds.

Finally, notice that

$$\left\| \frac{Q(\cdot \mid s, \theta) - Q(\cdot \mid s, 0)}{\theta} \right\| = \int_Y \left| \frac{q(y \mid s, \theta) - q(y \mid s, 0)}{\theta} \right| v(dy)$$

$$\leqslant \int_Y z(y, s) \, v(dy).$$

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

Proof of Corollary 3.　We need to show that the conditions of Theorem 1 are satisfied. Conditions b) and F1 imply that $F_\theta \in BV'([y, \bar{y}])$; since $BV'([y, \bar{y}])$ and $ca[y, \bar{y}]$ 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 $[y, \bar{y}]$, and let $\epsilon > 0$; given any $\{\theta_n\}$ that converges to zero, we must find an N such that for all $n \geq N$

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

The left side is equal to

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

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

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

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

Proof of Proposition 4.　Let $P(\cdot \mid y, \theta)$ denote a version of the posterior such that $P(\cdot \mid y, \theta_n) \xrightarrow{w} \mu(\cdot)$ for each y and for any sequence $\{\theta_n\}$ that converges to zero, and set $P(\cdot \mid y, 0) = \mu(\cdot)$ for all y. 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 $(a_n, \theta_n) \to (\bar{a}, 0)$, we have

$$\left| g(a_n, \theta_n, y) - g(\bar{a}, 0, y) \right|$$

$$= \left| \int_S u(a_n, s) \, P(ds \mid y, \theta_n) - \int_S u(\bar{a}, s) \, P(ds \mid y, 0) \right|$$

$$\leq \left| \int_S u(a_n, s) \, P(ds \mid y, \theta_n) - \int_S u(\bar{a}, s) \, P(ds \mid y, \theta_n) \right|$$

$$+ \left| \int_S u(\bar{a}, s) \, P(ds \mid y, \theta_n) - \int_S u(\bar{a}, s) \, P(ds \mid y, 0) \right|$$

$$\leqslant \int_S |u(a_n, s) - u(\bar{a}, s)| \, P(ds \mid y, \theta_n)$$

$$+ \left| \int_S u(\bar{a}, s) \, P(ds \mid y, \theta_n) - \int_S u(\bar{a}, s) \, P(ds \mid y, 0) \right|. \tag{9}$$

Since S and A are compact metric spaces and $u: A \times S \to R$ is continuous, $u(a_n, s) \to u(\bar{a}, s)$ uniformly (Dixmier [8, Theorem 6.1.13]) and therefore, given any $\epsilon > 0$, there is always an N_ϵ sufficiently large such that for all $n \geqslant N_\epsilon$

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

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 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 $B \in \mathscr{B}_S$.

Now, since A is convex and compact and $g(\cdot, \theta, y)$ is strictly concave, we have that for each y there is a unique solution $d^*(y, \theta)$ to problem (5). 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 $\bar{d} \in A$ and $\bar{d} \neq d^*(y, 0)$. Fix $a \in A$. For every k, we must of course have $g(d^*(y, \theta_{n_k}), \theta_{n_k}, y) \geqslant g(a, \theta_{n_k}, y)$. By the continuity of g on $A \times \{0\}$, we have $g(\bar{d}, 0) \geqslant g(a, 0)$. Since $a \in A$ was arbitrary, $\bar{d}$ must solve (5) for $\theta = 0$, contradicting uniqueness. Hence, $\lim_{\theta \to 0^+} d^*(y, \theta) = d^*(y, 0)$ for each y, and $d^*(y, 0)$ is independent of y.

The measurability of $d^*(y, \theta)$ follows from the fact that, for each θ, the conditions of the Measurable Maximum Theorem (Aliprantis and Border [1, Theorem 17.18]) are satisfied. ∎

Proof of Corollary 5. We only need show that the weak convergence condition holds. For any y that satisfies (ii)(a) and given any set $B \in \mathscr{B}_S$, 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 any y satisfying (ii)(b), define $P(B \mid y, \theta) = \mu(B)$ for all $\theta < \theta_y$. (For $\theta \geqslant \theta_y$, take any version of the posterior). Finally, for the ν-measure zero set that violates (ii) set $P(B \mid y, \theta) = \mu(B)$ for all $\theta \in \Theta$. This defines a version of the posterior for every $y \in Y$ and $\theta \in \Theta$.

To complete the argument that this version weakly converges to μ, we need only consider a y that satisfies (ii)(a). Take any $\theta_n \to 0$. (i)–(iii) yield that

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

where the application of the LDCT is justified by (i), and the last equality follows since $q(y \mid s, 0)$ does not depend on 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. ∎

REFERENCES

1. C. Aliprantis and K. Border, "Infinite Dimensional Analysis. A Hitchhiker's Guide," Springer-Verlag, Heidelberg, 1999.
2. K. Arrow, Informational structure of the firm, *Amer. Econ. Rev.* **75** (1985), 303–307.
3. D. Berstsekas and S. Shreve, "Stochastic Optimal Control: The Discrete-Time Case," Athena Scientific, Belmont, 1978.
4. P. Billingsley, "Probability and Measure," Wiley, New York, 1986.
5. P. Bolton and C. Harris, Strategic experimentation, *Econometrica* **67** (1999), 349–374.
6. D. Bradford and H. Kelejian, The value of information for crop forecasting in a market system: some theoretical issues, *Rev. Econ. Stud.* **44** (1977), 519–553.
7. A. Creane, Experimentation with heteroskedastic noise, *Econ. Theory* **4** (1994), 275–286.
8. J. Dixmier, "General Topology," Springer-Verlag, New York, 1984.
9. N. Dunford and J. Schwartz, "Linear Operators. Part I: General Theory," Wiley, New York, 1957.
10. D. Easley and N. Kiefer, Controlling a stochastic process with unknown parameters, *Econometrica* **56** (1988), 1045–1064.
11. X. Freixas and R. Kihlstrom, Risk aversion and information demand, *in* "Bayesian Models of Economic Theory" (M. Boyer and R. Kihlstrom, Eds.), pp. 93–104, Elsevier, Amsterdam, 1984.
12. S. Grossman, R. Kihlstrom, and L. Mirman, A Bayesian approach to the production of information and learning by doing, *Rev. Econ. Stud.* **44** (1977), 533–547.
13. P. Halmos, "Measure Theory," Springer-Verlag, New York, 1974.
14. J. Harrington, Experimentation and learning in a differentiated-products duopoly, *J. Econ. Theory* **66** (1995), 275–288.
15. N. Kiefer and Y. Nyarko, Optimal control of an unknown linear process with learning, *Int. Econ. Rev.* **30** (1989), 571–586.

16. R. Kihlstrom, A Bayesian model of demand for information about product quality, *Int. Econ. Rev.* **15** (1974), 99–118.

17. R. Kihlstrom, A general theory of demand for information about product quality, *J. Econ. Theory* **8** (1974), 413–439.

18. R. Kihlstrom, L. Mirman, and A. Postlewaite, Experimental consumption and the Rothschild effect, *in* "Bayesian Models of Economic Theory" (M. Boyer and R. Kihlstrom, Eds.), pp. 279–302, Elsevier, Amsterdam, 1984.

19. J. Laffont, On the welfare analysis of rational expectations equilibria with asymmetric information, *Econometrica* **53** (1985), 1–29.

20. T. Lewis and D. Sappington, Supplying information to facilitate price discrimination, *Int. Econ. Rev.* **35** (1994), 309–325.

21. A. McLennan, Price dispersion and incomplete learning in the long run, *J. Econ. Dynam. Control* **7** (1984), 331–347.

22. L. Mirman, L. Samuelson, and E. Schlee, Strategic information manipulation in duopolies, *J. Econ. Theory* **62** (1994), 363–384.

23. L. Mirman, L. Samuelson, and A. Urbano, Monopoly experimentation, *Int. Econ. Rev.* **34** (1993), 549–563.

24. G. Moscarini and L. Smith, Wald revisited: the optimal level of experimentation, *Econometrica* **69** (2001), 1629–1644.

25. G. Moscarini and L. Smith, The law of large demand for information, *Econometrica*, in press.

26. R. Radner, Uncertainty and general equilibrium, *in* "The New Palgrave: General Equilibrium" (J. Eatwell, M. Milgate, and P. Newman, Eds.), pp. 305–323, Norton, New York, 1989.

27. R. Radner and J. Stiglitz, A nonconcavity in the value of information, *in* "Bayesian Models of Economic Theory" (M. Boyer and R. Kihlstrom, Eds.), pp. 33–52, Elsevier, Amsterdam, 1984.

28. H. Royden, "Real Analysis," Macmillan, New York, 1988.

29. N. Singh, Monitoring and hierarchies: the marginal value of information in a principal–agent model, *J. Polit. Econ.* **93** (1985), 599–609.

30. J. Stiglitz, The contributions of the economics of information to twentieth century economics, *Quart. J. Econ.* **115** (2000), 1441–1478.

31. N. Stokey, R. Lucas, and E. Prescott, "Recursive Methods in Economic Dynamics," Harvard University Press, Cambridge, 1989.

32. I. Tonks, Bayesian learning and the optimal investment decision of the firm, *Econ. J.* **93** (1983), 86–97.

33. D. Trefler, The ignorant monopolist: optimal learning with endogenous information, *Int. Econ. Rev.* **34** (1993), 565–581.

34. R. Wilson, Informational economies of scale, *Bell J. Econ.* **6** (1975), 184–195.