Econometrica, Vol. 70, No. 6 (November, 2002), 2351–2366

THE LAW OF LARGE DEMAND FOR INFORMATION

BY GIUSEPPE MOSCARINI AND LONES SMITH[1]

An unresolved problem in Bayesian decision theory is how to value and price information. This paper resolves both problems assuming inexpensive information. Building on Large Deviation Theory, we produce a generically complete asymptotic order on samples of i.i.d. signals in finite-state, finite-action models. Computing the marginal value of an additional signal, we find it is eventually exponentially falling in quantity, and higher for lower quality signals. We provide a precise formula for the information demand, valid at low prices: asymptotically a constant times the log price, and falling in the signal quality for a given price.

KEYWORDS: Demand for information, logarithmic demand, value of information, Bayesian decision theory, comparison of experiments, large deviation theory.

1. INTRODUCTION

AN UNRESOLVED PROBLEM in Bayesian decision theory is how to value and price information. For instance, Blackwell's Theorem asserts that the value of different informative statistical experiments (signals) is generically incomparable across all decision makers—sufficiency being an extremely partial order.[2] The theory of information demand is also problematic: Under smoothness assumptions on information and payoff functions, Radner and Stiglitz (1984) argued that the marginal value of information is initially zero. Consequently, the value of information is not globally concave, and first order conditions alone do not describe demand.[3] So for those in the business of buying or selling information, economists have little to say.

This paper attempts to fill this gap, for the case of inexpensive information units. One economic motivation of this assumption owes to the rise of the Internet. Web-based distribution of information—be it databases, encyclopedias,

[1] This revision reflects the comments of a co-editor and three referees on our April 2000 version. We have also benefited from the feedback of seminar audiences at Princeton, NYU, and Yale, and specifically thank Jon Levin for detailed early comments, and Ed Schlee for drawing our attention to the work of Torgersen. Of course, any mistakes are our responsibility alone. Giuseppe gratefully acknowledges the support of the Alfred P. Sloan Foundation, a Yale Junior Faculty Fellowship, and the hospitality of the Department of Economics of New York University while working on this project. Lones acknowledges the support of the NSF.

[2] In response to this difficulty, Athey and Levin (2001) develop a complete ordering by simultaneously restricting the class of payoff functions and signal distributions. Persico (2000) shows the usefulness of such orderings in auction theory. Both papers build on Lehmann (1988).

[3] Chade and Schlee (2001) clarify the nature of the assumptions underlying Radner and Stiglitz (1984).

or news services—has dramatically reduced its marginal cost.[4] While economists have focused on the valuation and pricing of a single informative signal, these technological changes speak to the importance of the large demand special case. Alternatively, the information acquired is 'large' whenever the cost of an extra sample is very small *relative* to the payoff stakes—as with tests of new potentially dangerous devices, like motor vehicles, aircraft, or space vehicles, or in the final phase of clinical trials of new drugs.

Throughout this paper, we interpret the quantity of information as the number of signal draws purchased nonsequentially.[5] With this natural definition, we find that for large samples there exists a generically complete order for statistical experiments with finitely many states and actions. This is true despite Blackwell's restrictive order for each signal draw. And despite the negative message of Radner and Stiglitz for small quantities of information, we exhibit a falling demand curve for low enough prices and hence large demands; we then derive its asymptotic formula, which is logarithmic in price.

In statistical decision models, as evidence accumulates the truth gets revealed by the Strong Law of Large Numbers (SLLN), and one eventually makes the right decision. So decision making is about avoiding mistakes, which in turn are deviations from the SLLN. We take inspiration from Chernoff's (1952) fundamental connection between simple statistical hypotheses tests and Large Deviation Theory. We refine this logic for Bayesian decision problems characterizing the marginal value of information, and thereby the demand for informative samples; finally, we also extend the theory to the multi-state case.

Specifically, Chernoff employed Cramér's (1938) asymptotic expansion for the chance of a large deviation of a sample *mean* from its SLLN limit. In a Bayesian world, rational behavior in the two state case depends solely on the total log-likelihood ratio (logLR) of the observed sample—or the *sum* of i.i.d. random variables. We extend Cramér's result to compute the chance of large deviations of such running sample sums (Lemma 1).

We find that for any (imperfectly) informative signal σ, a unique *efficiency index* $\rho \in (0, 1)$ exists such that the value of perfect information less that of n independent draws from σ behaves as ρ^n for large n (Theorem 1). Therefore, when one may sample from a large pool of conditionally independent draws of the signal σ, the scalar $1/\rho$ measures the value of σ and yields a *generically complete* order over signals for any decision maker.

We then extend the order on the total value of information to the margin, and thereby demand. We find that the marginal value of the nth independent draw behaves as ρ^n, for large n, and so is eventually falling monotonically (Theorem 2). This yields our 'Law of Large Demand' for information (Theorem 3): For all low enough prices $p > 0$, the demand for information is asymptotically $(\log p)/\log \rho$,

[4] See Shapiro and Varian (1999). For a concrete example, the Roper Center (a global seller of information services, such as opinion polls) charges per question fee of \$1 to academics, \$1.50 to others.

[5] By contrast, Moscarini and Smith (2001) show that in a *dynamic* continuous time world, one-shot nonsequential sampling is still optimal given discounting and a constant marginal cost of information.

and thus falling in the signal quality $1/\rho$. In fact, at small prices, we ascertain the information demand within an integer. More specifically, information demand lies within one of $[\log p + \frac{1}{2}\log(-\log p) + (\text{constant})]/\log\rho$, where the constant depends on preferences and the signal.

Our logarithmic demand formula should have many applications; for instance, it yields boundary behavior of the information demand curve in a general equilibrium setting. While this formula obtains for large demand, we hope that in applied analyses of information provision, a logarithmic form offers some guidance as a benchmark; e.g. it invalidates alternative functional forms, such as isoelastic. It also implies that more valuable signals have a lower marginal value schedule, and so are less demanded.

Section 5 provides a key inequality establishing the robustness of our results to finitely many states and actions (Theorem 4). In particular, the inference problem for large samples with many states of nature is dominated by the two states hardest to distinguish; this pair determines the signal's quality measure ρ. A Bayesian decision-maker eventually optimally focuses on a worst-case scenario independently of her beliefs and preferences.

2. THE MODEL

A Decision Maker ($\mathcal{DM}$) must choose an action a from a finite menu $\mathbb{A} = \{a_1, \ldots, a_K\}$. The $\mathcal{DM}$ has a full support prior probability density $q(\theta)$ on state $\theta \in \Theta = \{\theta_1, \ldots, \theta_M\}$. Action a yields the $\mathcal{DM}$ vNM utility $u(a, \theta)$ in state θ, where $u : \mathbb{A} \times \Theta \to \mathbb{R}$. No action is dominated (which is without loss of generality), and there is a unique best action $a^*(\theta) = \arg\max_{a\in\mathbb{A}} u(a, \theta)$ in each state θ (as is generically true).

Let $\mathcal{E} = \langle f(\cdot|\theta), \theta \in \Theta\rangle$ be an *experiment* (sometimes called a *signal*). By this, we mean a family of state-dependent probability densities on outcomes in $\mathbb{X}$, each associated to a probability measure μ^θ on a measurable space $(\mathbb{X}, \mathcal{F})$. So given the true state θ, $\mathcal{E}$ can be represented by a random variable X with the density $f(\cdot|\theta)$ over outcomes in $\mathbb{X}$. Each signal outcome is assumed imperfectly informative, $f(\cdot|\theta)$ having full support on $\mathbb{X}$.

Before choosing an action, the $\mathcal{DM}$ chooses a sample size n, and then observes the outcome of an n-sample $X^n = (X_1, \ldots, X_n) \in \mathbb{X}^n$ of experiments $\mathcal{E}$. Observations are assumed to be independent, conditional on the state.[6] After seeing the joint realization $X^n = x^n$, the $\mathcal{DM}$ then updates her prior beliefs to the posterior $\Pr(\theta|x^n, q)$, and takes the action that maximizes expected utility given the sample, namely $a^*(x^n) = \arg\max_{a\in\mathbb{A}} E_\theta[u(a, \theta)|x^n, q]$. The $\mathcal{DM}$'s ex ante expected payoff $V_{q,u}(n)$ from sampling n observations is her maximum expected utility,

[6] But in fact, the analysis obtains provided the observation process is *exchangeable* (i.e. the probability of an infinite sequence is independent of the permutation), which is quite natural for repeated experiments. For then, by De Finetti's Representation Theorem (Theorem 1.48 in Schervisch (1995)), the observations are conditionally iid, after appropriately re-defining the state of nature.

depending on her prior and utility function:

$$V_{q,u}(n) = E_{X^n}\big[E_\theta[u(a^*(X^n),\theta)|X^n,q]\big]$$
$$= \sum_{\theta\in\Theta} \frac{\int_{\mathbb{X}^n} u(a^*(x^n),\theta)\prod_{i=1}^n f(x_i|\theta)\,dx^n}{\sum_{\tau\in\Theta}\int_{\mathbb{X}^n}\prod_{i=1}^n f(x_i|\tau)q(\tau)\,dx^n} q(\theta),$$

where $\prod_{i=1}^n f(x_i|\theta)q(\theta)$ is the joint probability density of the state θ and observations x^n. We suppress the subscripts u or q of the payoff function V, when there is no ambiguity. As the $\mathcal{DM}$ may always ignore observations, $V(n)$ is nondecreasing in the sample size n.

For pedagogical reasons, the paper focuses until Section 5 on the binary-state, binary-action case with $\Theta = \{L, H\}$ and $\mathbb{A} = \{A, B\}$, as in a classical test of simple hypotheses. Absent a dominated action, we assume without loss of generality that action B is best in state H, so that $u(A,L) > u(B,L)$ and $u(B,H) > u(A,H)$. In this context, denote the belief $q = \Pr(H)$, so that $V_q(0)$ is a piecewise linear, convex function of q. The $\mathcal{DM}$ then optimally chooses action B iff $q \geq \hat{q}$, for some $\hat{q} \in (0,1)$.

3. LARGE DEVIATIONS AND THE VALUE OF INFORMATION

A. *Expected Payoffs via Error Chances*

The $\mathcal{DM}$ selects A if and only if her belief falls below a cutoff:

$$q_n = \Pr(H|X^n) < \hat{q} = \frac{u(A,L)-u(B,L)}{u(B,H)-u(A,H)+u(A,L)-u(B,L)} \in (0,1).$$

The expected payoff is then $\Pr(q_n < \hat{q}|L)u(A,L) + \Pr(q_n \geq \hat{q}|L)u(B,L)$ in state L, and $\Pr(q_n \geq \hat{q}|H)u(B,H) + \Pr(q_n < \hat{q}|H)u(A,H)$ in state H. Let the two error probabilities be $\alpha_n = \Pr(q_n \geq \hat{q}|L)$ and $\beta_n = \Pr(q_n < \hat{q}|H)$. Then

$$V_{q,u}(n) = (1-q)[(1-\alpha_n)u(A,L) + \alpha_n u(B,L)]$$
$$+ q[(1-\beta_n)u(B,H) + \beta_n u(A,H)].$$

So if the *full-information payoff* $V^*_{q,u} = (1-q)u(A,L) + qu(B,H)$ is the expected payoff given an ex post optimal action, then the *Full Information Gap* (FIG) equals

$$V^*_{q,u} - V_{q,u}(n) = \alpha_n(1-q)[u(A,L)-u(B,L)] \tag{1}$$
$$+ \beta_n q[u(B,H)-u(A,H)]$$
$$\equiv \alpha_n w^L(q,u) + \beta_n w^H(q,u)$$

for positive constants $w^\theta(q,u)$. Since $f(x^n|\theta) = \prod_{i=1}^n f(x_i|\theta)$ by conditional independence, the sample *log likelihood ratio* (log-LR) of states L to H is $S_n^L = \sum_{i=1}^n \ell^L(X_i)$, where $\ell^L(x_i) = \log[f(x_i|L)/f(x_i|H)]$. Analogously define the log-LR $\ell^H(X_i) = -\ell^L(X_i)$, and its sum $S_n^H = -S_n^L$. Write the posterior belief in log-likelihood form:

$$q_n = \frac{qf(X^n|H)}{qf(X^n|H) + (1-q)f(X^n|L)} = \frac{1}{1 + e^{\log[(1-q)/q] + S_n^L}}.$$

The optimality rule $q_n \leq \hat{q}$ is thus the same as $S_n^L > \log[(1-\hat{q})/\hat{q}] - \log[(1-q)/q] \equiv \xi(q, u)$. Hence:

$$\text{(2)} \qquad \alpha_n = \Pr\big(S_n^L \leq \xi(q,u)|L\big) \quad \text{and} \quad \beta_n = \Pr\big(S_n^H < -\xi(q,u)|H\big).$$

B. *Asymptotic Tests*

Perhaps the log-LR cannot cross the threshold $\xi(q, u)$ until at least $N_0 > 1$ draws, so that small samples $n \leq N_0$ are worthless. Define $\lambda^\theta \equiv E[\ell^\theta(X_i)|\theta] > 0$ for $\theta = L, H$.[7] Then a large enough sample size must have positive value, since in state L (say) $\lim_{n\to\infty} S_n^\theta/n = \lambda^\theta > 0$ (a.s.) by the SLLN. So we eventually a.s. have $S_n^\theta > \xi(q, u)$. The dependence of the error chances α_n, β_n—and thus the payoff $V(n)$—on the sample size n reduces not to the Law of Large Numbers, but instead to the Theory of Large Deviations for sums of independent identically distributed random variables, which studies the chance of tail events. Since $\lim_{n\to\infty} S_n^\theta/n - \lambda^\theta = 0 > -\lambda^\theta$ in states $\theta = L, H$, given $\ell_i^H = -\ell_i^L$ and $\lambda^H = E[\ell_i^H|H] > 0$,

$$\alpha_n = \Pr\left(\frac{S_n^L}{n} - \lambda^L \leq \frac{\xi(q,u)}{n} - \lambda^L|L\right) \quad \text{and}$$
$$\beta_n = \Pr\left(\frac{S_n^H}{n} - \lambda^H < \frac{-\xi(q,u)}{n} - \lambda^H|H\right).$$

To characterize the chances of large deviations, we make some technical assumptions about random variables Y, such as the log-LR $\ell^\theta(X_i)$. First, Y has a finite mean $E[Y]$ and variance $V[Y] = \sigma^2$, with $E[Y] > 0$ and $\Pr(Y < 0) > 0$. Second, Y is not a *lattice random variable*, i.e. with range $\{y_0, y_0 \pm d, y_0 \pm 2d, \dots\}$ for some d, y_0. Third, $E[e^{\bar{t}|Y|}] < \infty$ for some $\bar{t} \neq 0$.[8] Let $\mathcal{M}(t) \equiv E[e^{tY}]$ be the moment generating function (MGF) of Y, and define $\rho = \inf_t \mathcal{M}(t) \equiv \mathcal{M}(\tau) < 1$. Let $S_n = Y_1 + \cdots + Y_n$. Chernoff exploited Cramér's (1938) asymptotic formula

[7] By Proposition 2.92, Schervisch (1995), the integral

$$\lambda^L \equiv \int_x f(x|L) \log[f(x|L)/f(x|H)]\, dx$$

is nonnegative, and strictly so if the densities $f(\cdot|H)$ and $f(\cdot|L)$ differ.

[8] This requires that all moments of Y be finite; the standard SLLN requires only four finite moments.

for the chance of a large deviation of a sample mean:

$$\text{(3)}\qquad \Pr\left(\frac{S_n}{n}<0\right)=\frac{\rho^n}{\gamma\sqrt{2\pi n}}\left(1+\sum_{i=1}^{k}\frac{b_i}{n^i}\right)\left(1+O\left(\frac{1}{n^{k+1}}\right)\right)$$
$$=\frac{\rho^n}{\gamma\sqrt{2\pi n}}\left(1+O\left(\frac{1}{n}\right)\right)$$

for some sequence $\{b_i\}_{i=1}^k$, for every $k=1,2,\dots,$ and $\gamma\equiv\sigma|\tau|$.

This theory, applied to our log-likelihood ratio process $Y=\ell^\theta(X_i)$, yields a special structure. To see this, define the *Hellinger transform* of an experiment $\mathscr{E}$:

$$\mathscr{H}_{\mathscr{E}}(t)\equiv\int_{\mathbb{X}}[f(x|L)]^t[f(x|H)]^{1-t}\,dx=\int_{\mathbb{X}}f(x|H)e^{t\ell^L(x)}\,dx.$$

This is the moment generating function (MGF) of the log-LR of either state L versus H in state H, or vice versa, after change of variables: respectively, $\mathscr{M}^H(t)\equiv\mathscr{H}_{\mathscr{E}}(-t)$ and $\mathscr{M}^L(t)\equiv\mathscr{H}_{\mathscr{E}}(t+1)$. Cramér's theory suggests considering the minimum of this MGF. Note that when $\mathscr{E}$ is *informative*, namely $f(\cdot|H)$ and $f(\cdot|H)$ differ, such a minimum uniquely exists: Indeed, $\mathscr{H}_{\mathscr{E}}(t)$ is a strictly convex function on $(0,1)$ obeying $\mathscr{H}_{\mathscr{E}}(0)=\mathscr{H}_{\mathscr{E}}(1)=1$ with slopes $\mathscr{H}'_{\mathscr{E}}(0)=-\lambda^L<0$ and $\mathscr{H}'_{\mathscr{E}}(1)=\lambda^H>0$. Put

$$\rho_{\mathscr{E}}\equiv\min_{\mathbb{R}}\mathscr{H}_{\mathscr{E}}(t)=\mathscr{H}_{\mathscr{E}}(\tau_{\mathscr{E}})\quad\text{at some }\tau_{\mathscr{E}}\in(0,1).$$

Remarkably, the quantity $\rho_{\mathscr{E}}$ is state-independent, since $\rho_{\mathscr{E}}=\mathscr{M}^L(\tau^L)=\mathscr{M}^H(\tau^H)$, where τ^θ is the minimizer of the MGF of ℓ^θ, and $\tau_{\mathscr{E}}=\tau^L+1=-\tau^H\in(0,1)$. So we can treat $\rho_{\mathscr{E}}$ as the unique *efficiency index* of the experiment, measuring how easy large samples can distinguish $f(\cdot|H)$ and $f(\cdot|L)$. Indeed, $\rho_{\mathscr{E}}$ has all the right properties for measuring an experiment's informativeness in large samples. For instance, $\mathscr{H}_{\mathscr{E}_1\times\mathscr{E}_2}(t)=\mathscr{H}_{\mathscr{E}_1}(t)\mathscr{H}_{\mathscr{E}_2}(t)$ for the joint experiment $\mathscr{E}_1\times\mathscr{E}_2$, implying $\rho_{\mathscr{E}_1\times\mathscr{E}_2}=\rho_{\mathscr{E}_1}\rho_{\mathscr{E}_2}$. (See Torgersen (1991, §1.4).)

Chernoff (1952) used $\lim_{n\to\infty}[\log\Pr(S_n/n\le\kappa|\theta)]/n=\log\rho_{\mathscr{E}}$ for $\theta=H,L$, and any $\kappa<\lambda^\theta$, to rank asymptotically classical hypothesis tests based on a sample *mean*. This formula followed from the lead term of Cramér's expansion (3). By contrast, our Bayesian error probabilities α_n,β_n are chances of large deviations of sample *sums*. By exploiting the first two terms of (3), we now extend Chernoff's finding to our Bayesian framework.

LEMMA 1: *Let $\mathscr{M}^\theta(t)$ be the MGF of $\ell^\theta(X)$ in state θ with minimum $\rho_{\mathscr{E}}$ achieved at $\tau^\theta\equiv\arg\min\mathscr{M}^\theta(t)<0$. Put $\gamma^\theta\equiv\sigma|\tau^\theta|$. Then the error chances α_n and β_n satisfy*

$$\text{(4)}\qquad \alpha_n=\frac{e^{|\tau^L|\xi(q,u)}\rho_{\mathscr{E}}^n}{\gamma^L\sqrt{2\pi n}}\left(1+O\left(\frac{1}{\sqrt{n}}\right)\right)\quad\textit{and}$$
$$\beta_n=\frac{e^{-|\tau^H|\xi(q,u)}\rho_{\mathscr{E}}^n}{\gamma^H\sqrt{2\pi n}}\left(1+O\left(\frac{1}{\sqrt{n}}\right)\right).$$

C. *The Asymptotic Value of Information*

Since the Full Information Gap (1) is a weighted average of the error chances α_n and β_n, each of the form in (4), the gap vanishes as fast as do α_n and β_n. The next theorem is then immediate in light of (1)–(4), given the following constant:

$$c \equiv \frac{w^L(q,u)e^{|\tau^L|\xi(q,u)}}{\gamma^L\sqrt{2\pi}} + \frac{w^H(q,u)e^{-|\tau^H|\xi(q,u)}}{\gamma^H\sqrt{2\pi}}.$$

THEOREM 1: *Given a $\mathcal{DM}(q,u)$ and an experiment $\mathscr{E}$ with efficiency index $\rho_{\mathscr{E}}$, the full Information Gap obeys*

$$FIG(n) \equiv V^*_{q,u} - V_{q,u}(n) = \frac{c\rho^n_{\mathscr{E}}}{\sqrt{n}}\left(1+O\left(\frac{1}{\sqrt{n}}\right)\right).$$

COROLLARY: *Fix experiments $\mathscr{E}_1$ and $\mathscr{E}_2$ with $\rho_{\mathscr{E}_1} < \rho_{\mathscr{E}_2}$. For generic (q,u), there exists $N < \infty$ such that for all $n \geq N$, an n-sample from $\mathscr{E}_1$ is strictly preferred to one from $\mathscr{E}_2$.*

In other words, *for any finite set of $\mathcal{DM}$'s, there exists $N < \infty$ such that for all larger sample sizes $n > N$, all $\mathcal{DM}$'s agree on the ordering of n-samples from $\mathscr{E}_1$ vs. $\mathscr{E}_2$*. One cannot rank different experiments with the same efficiency index ρ.[9] Viewing minimizing expected losses as maximizing expected utility, the corollary follows from Theorem 3.9 in Torgersen (1981) (hereafter T81) (which is weaker than our Lemma 1). Our novel results begin next with the marginal analysis, for which T81's Theorem 3.9 is too weak to build upon.[10]

The error term in Theorem 1 plays a central role in justifying our asymptotic analysis. For consider an approximate $\widehat{FIG}(n) \equiv c\rho^n_{\mathscr{E}}/\sqrt{n}$, omitting the variable $\mathcal{DM}$-idiosyncratic error terms. Ranking experiments by their approximate $\widehat{FIG}$s is based *only* on our efficiency index $\rho^n_{\mathscr{E}}$, for any $\mathcal{DM}$. While both the approximate $\widehat{FIG}(n)$ and the true $FIG(n)$ vanish as n explodes, the percentage divergence between the two vanishes too, being $O(1/\sqrt{n})$; thus, $\widehat{FIG}(n)$ and $FIG(n)$ approach each other much faster than either of them approaches zero. This will also hold true for the marginal value of information.

We briefly relate Theorem 1 to a result much more familiar among economists—Blackwell's (1953) Theorem. If $\mathscr{E}_1$ is statistically sufficient for $\mathscr{E}_2$, then any n-sample from $\mathscr{E}_1$ is sufficient for that from $\mathscr{E}_2$, and thus $\rho_{\mathscr{E}_1} \leq \rho_{\mathscr{E}_2}$, simply because $\mathscr{H}_{\mathscr{E}_1}(t) \leq \mathscr{H}_{\mathscr{E}_2}(t)$ at all t (see p. 358 of Torgersen (1991)). But if $\mathscr{E}_1$ is not sufficient for $\mathscr{E}_2$, nor conversely, the same is true of their n-replicas, and Blackwell's Theorem has nothing to say. Our Theorem 1 gives a generically

[9] For a recent helpful contrast, Sandroni (2000) asks which trader dominates in the long run when prior beliefs differ. He conditions on the true state θ, and so may rank traders by their λ^θ measures. We must take an unconditional approach, and yet still find a unique scalar index ρ.

[10] Unlike our error computation, T81 essentially finds that $FIG = \rho^n f(n)$, where $(1/n)\log f(n) = o(1)$.

complete asymptotic ranking that favors the experiment with the smaller ρ. In Section 5, we show that this is true for any finite number of actions and states. While the conclusion obtains for all priors q and payoffs u, the threshold sample size N depends not only on the signal structure $\langle \Theta, f \rangle$ like Blackwell's Theorem, but also on the decision problem, namely the belief q and payoffs u. So no finite n-sample of independent signals from $\mathscr{E}_1$ is ever sufficient for an n-sample from $\mathscr{E}_2$.

D. *The Asymptotic Marginal Value of Information*

To ascertain not only the $\mathscr{DM}$'s value of information but also his demand, we now consider the asymptotic incremental value of the nth signal draw; for this, we introduce the notation $\Delta g(n) \equiv g(n+1) - g(n)$, for any function $g(n)$. We investigate the asymptotic properties of the increment in the chance of a large deviation for a unit increment in the sample size.

Lemma 1 allows us to deduce an asymptotic ordering of the marginal value of information, and thereby an asymptotically falling marginal value of information.

THEOREM 2: *Fix an experiment $\mathscr{E}$ and a $\mathscr{DM}(q,u)$. The marginal value of the* nth *sample from $\mathscr{E}$ vanishes as does $\rho_{\mathscr{E}}^n/\sqrt{n}$*:

$$\Delta V(n) = c(1-\rho_{\mathscr{E}})\frac{\rho_{\mathscr{E}}^n}{\sqrt{n}}\left(1 + O\left(\frac{1}{\sqrt{n}}\right)\right).$$

Moreover, there exists $N < \infty$ such that $\Delta V(n) > \Delta V(n+1) > 0$ for all $n \geq N$.

PROOF: Taking first differences of the value yields

$$\begin{aligned}\Delta V(n) &= FIG(n) - FIG(n+1)\\ &= c\frac{\rho_{\mathscr{E}}^n}{\sqrt{n}}\left(1 + O\left(\frac{1}{\sqrt{n}}\right)\right) - c\frac{\rho_{\mathscr{E}}^{n+1}}{\sqrt{n+1}}\left(1 + O\left(\frac{1}{\sqrt{n+1}}\right)\right).\end{aligned}$$

The first claim follows because $\sqrt{n/(n+1)} \to 1$. For the second claim, take second differences:

$$\Delta^2 V(n) = \Delta V(n+1) - \Delta V(n) = -c(1-\rho_{\mathscr{E}})^2\frac{\rho_{\mathscr{E}}^n}{\sqrt{n}}\left(1 + O\left(\frac{1}{\sqrt{n}}\right)\right).$$

By the last two equations, for N large enough, $\Delta V(n) > 0 > \Delta^2 V(n)$ for all $n \geq N$. *Q.E.D.*

So by Theorems 1 and 2, as the quality of information grows, i.e. the efficiency index $\rho_{\mathscr{E}}$ falls, the expected value of an n-sample rises, as expected, while less obviously so, the marginal value of the nth observation falls.

For a revealing comparison with the statistical literature, Mammen (1986) studies the behavior of the *deficiency* of n with respect to $n+1$ signal draws—namely, the most one can gain by using $n+1$ versus n observations, where the maximum is taken over priors q and payoffs u in $[-1, 1]$ (see Le Cam (1964)). He finds that this marginal deficiency is of order $O(1/n)$. Loosely, as a "worst-case analysis" across priors and payoffs, the deficiency is rather slowly vanishing. By contrast, Full Information Gaps vanish exponentially fast, since the $\mathcal{DM}$'s prior and payoffs do not vary with the sample size n.

4. THE DEMAND FOR INFORMATION

Suppose that the $\mathcal{DM}$ faces a constant price $p > 0$ for each independent observation of $\mathcal{E}$. Assume that he chooses n to maximize $V(n) - pn$. His information demand is the number of signals $n(p) \in \{1, 2, \dots\}$ that he buys. With a smooth concave value, demand solves $V'(n(p)) = p$, and so if the marginal value $V'(n)$ vanishes exponentially, then demand is logarithmic. Of course, demand is discrete, and perhaps poorly behaved; also our approximate marginal value contains both exponential and geometric factors. Still, we now prove that this intuitive log demand formula is almost precise, with an extra term owing to the geometric factor, and we very tightly bound the approximation error. We underscore that, at small prices, our demand formula is exact *to the nearest integer*.

THEOREM 3: *Fix an experiment $\mathcal{E}$ with efficiency index $\rho_{\mathcal{E}}$. For almost all p, demand $n(p)$ is single-valued, and decreasing in p. Also, there exists $\bar{p} > 0$ such that for all $p \in (0, \bar{p})$, demand is within 1 of*

$$\frac{\log p + \frac{1}{2}\log[(\log p)/\log \rho_{\mathcal{E}}] - \log[c(1-\rho_{\mathcal{E}})]}{\log \rho_{\mathcal{E}}}.$$

So for small p: $n(p) \sim (\log p)/(\log \rho_{\mathcal{E}})$; further, $n(p)$ monotonically rises in $\rho_{\mathcal{E}}$, at fixed p.

PROOF: In the Appendix, we prove that for almost all p, there is a unique demand[11] $n(p)$, with $\lim_{p\to 0} n(p) = \infty$. Furthermore, optimality clearly demands that any optimal demand $n(p)$ solves the "discrete FOC"

$$\Delta V(n(p)) \le p \le \Delta V(n(p) - 1). \tag{5}$$

1. *Lead Log Term:* Let $C_1 \equiv 1/\log \rho_{\mathcal{E}}$. Take logs of (5) using $\Delta V(n)$ from Theorem 2:

$$\begin{aligned} n(p)/C_1 - \frac{1}{2}\log n(p) + Q(n(p)) \\ \le \log p \le [n(p)-1]/C_1 - \frac{1}{2}\log(n(p)-1) + Q(n(p)-1) \end{aligned} \tag{6}$$

[11] Indeed, when nonunique, the demand must be two consecutive integers. We owe this fine point and its quick proof to an alert referee.

where[12]

$$Q(n) \equiv \log c(1-\rho_{\mathscr{E}}) + \log(1+O(1/\sqrt{n})) \\ = (1+o_p(1))\log[c(1-\rho_{\mathscr{E}})] \equiv (1+o_p(1))C_2,$$

since $n(0)=\infty$. Multiply (6) by $C_1<0$, and use $(\log n)/n$ vanishing in n (so $\log n = o_p(1)n$):

$$(n(p)-1)(1+o_p(1)) + (1+o_p(1))C_1C_2 \\ \leq C_1\log p \leq n(p)(1+o_p(1)) + (1+o_p(1))C_1C_2 \\ \Longrightarrow C_1(\log p)(1+o_p(1)) \leq n(p) \leq C_1(\log p)(1+o_p(1))$$

as $C_1(\log p)o_p(1)$ absorbs all constants. So $n(p) = (1+\delta_{n(p)})C_1(\log p)$, where $\delta_{n(p)} = o_p(1)$.

2. *Log-Log Term:* We have $\log n(p) = \log(-\log p)(1+o_p(1))$, and so also $\log[n(p)-1] = \log([1+o_p(1)]n(p)) = \log(-\log p)(1+o_p(1))$. To quantify $\delta_{n(p)}$, substitute the expressions for $n(p)$, $\log n(p)$, and $\log[n(p)-1]$ into (6). If we then subtract $n(p)\log\rho_{\mathscr{E}} = \log p$, and absorb $Q(n)$, $Q(n-1)\approx C_2$ into $\log(-\log p)(1+o_p(1))$:

$$\delta_{n(p)}\log p - \frac{1}{2}\log(-\log p)(1+o_p(1)) \\ \leq 0 \leq \delta_{n(p)}\log p - \frac{1}{C_1} - \frac{1}{2}\log(-\log p)(1+o_p(1)).$$

Similarly absorb $1/C_1$, divide by $\log p$, and use $\log n(p)/\log p = O((\log(-\log p))/\log p) = o_p(1)$, to get

$$\delta_{n(p)} = \frac{\log(-\log p)}{2\log p}(1+o_p(1)).$$

3. *Error Term:* Substitute $n(p) = C_1[\log p + (1/2)\log(-\log p)] + \phi(p)$ into (6), subtract $\log p$ globally, and rearrange terms, using $\log n(p) - \log(C_1\log p) = \log[1+o_p(1)] = o_p(1)$:

$$2\phi(p) + C_1\log(-C_1) + 2C_1C_2 + o_p(1) \\ \geq 0 \geq 2\phi(p) - 2 + C_1\log(-C_1) + 2C_1C_2 + o_p(1).$$

So all ϕ in the limit support of $\phi(p)$ (as $p\to 0$) obey $2\phi + C_1\log(-C_1) + 2C_1C_2 \in [0,2]$. That is, for all $\varepsilon>0$, there is $\bar{p}>0$, so that for all $p\in(0,\bar{p})$, we have the error bounds $-\varepsilon < \phi(p) + (\frac{1}{2}\log(-\log\rho_{\mathscr{E}}) + \log[c(1-\rho_{\mathscr{E}})])/(\log\rho_{\mathscr{E}}) < 1+\varepsilon$. *Q.E.D.*

[12] Here, $o_p(1)$ denotes any term vanishing as $p\to 0$, corresponding to terms $o(1)$ vanishing as $n\to\infty$.

5. MULTIPLE ACTIONS AND STATES

We now extend our main results (Theorems 1–3) to the more general case of finite M states and K actions. In fact, we need only extend Theorem 1, because its Corollary and the proofs of Theorems 2–3 solely rely on it and not on M or K. In Theorem 1, the constant c now becomes c_{MK}, reflecting all payoffs u_{ij} ($i = 1, 2, \dots, M$ and $j = 1, 2, \dots, K$).

A. *Two States, Multiple Actions*

First assume $K > 2$ undominated actions and two states. The optimal decision rule then entails a partition of $[0, 1]$ into K intervals bounded by $K - 1$ cutoff posterior beliefs, or equivalently, $K - 1$ cutoffs for the sample log-LR S_n^θ, say $\xi_2, \dots, \xi_K$. Just as with $K = 2$, write the value of an n-sample as a linear combination of error chances, $\alpha_{nj}^\theta \equiv \Pr(\xi_j < S_n^\theta \leq \xi_{j+1}|\theta)$ for $j = 1, 2, \dots, K$ (with $\xi_1 = -\infty$). From Lemma 1, these chances are of the same order as $\Pr(S_n^\theta \leq 0|\theta)$:

$$\begin{aligned}\alpha_{nj}^\theta &= \Pr\big(S_n^\theta \leq \xi_{j+1}|\theta\big) - \Pr\big(S_n^\theta \leq \xi_j|\theta\big)\\ &= \big(e^{|\tau^\theta|\xi_{j+1}} - e^{-|\tau^\theta|\xi_j}\big)\Pr\big(S_n^\theta \leq 0|\theta\big)(1 + O(1/\sqrt{n})).\end{aligned}$$

So the key properties of $\Pr(S_n^\theta \leq 0|\theta)$ carry over to α_{nj}^θ, and Theorem 1 holds with the multiplicative constant c modified to c_{2K}, which absorbs factors like $(e^{|\tau^\theta|\xi_{j+1}} - e^{-|\tau^\theta|\xi_j})$.

B. *Multiple States and Actions*

By merging states, assume $M > 2$ statistically distinguishable states $\Theta = \{\theta_1, \dots, \theta_M\}$, with one action dominant in each state (generically true), so $K \geq M$. Label actions so that a_i is best in θ_i: $u_{ii} > u_{ji}$ for $j \neq i$, where $u_{ji} \equiv u(a_j, \theta_i)$. *Insurance actions* $a_k (M < k \leq K)$ are best at unfocused beliefs.

Fix $\mathscr{E}$. For states $i, j = 1, 2, \dots, M$, $i \neq j$, define the 'pairwise' efficiency index of $\mathscr{E}$:

$$\rho_{ij} \equiv \min_{t\in[0,1]} \int f(x|\theta_i)^t f(x|\theta_j)^{1-t}\, dx,$$

suppressing a further $\mathscr{E}$ subscript on ρ. Label states so that $\rho_{12} = \max_{i, j\neq i} \rho_{ij}$.

Let $V(n)$ denote the expected value of an n-sample, and $V^* \equiv \sum_i q_i u_{ii}$ the full information value. Let $V_{ij}(n)$ be the value of n observations in the *ij-subdichotomy*—namely, where the $\mathscr{DM}$ additionally knows that $\theta \in \{\theta_i, \theta_j\}$. This yields updated prior beliefs $q_h/(q_i + q_j)$ on state $\theta_h = \theta_i, \theta_j$, but still any action a_k can be taken. Let V_{ij}^* be the full information payoff of the ij-dichotomy. In Appendix C, we prove the central result:

THEOREM 4: *Assume a unique* $\arg\max$ $ij = 12$ *of* ρ_{ij}. *Then there exists* $\bar\rho_{\mathscr{E}} < 1$ *such that*, *for all payoffs* u *and nondegenerate prior beliefs* q:

$$V^* - V(n) = (q_1 + q_2)[V_{12}^* - V_{12}(n)](1 + O(\bar\rho_{\mathscr{E}}^n)).$$

To wit, the FIG tends exponentially fast to the total chance q_1+q_2 of states $\{\theta_1, \theta_2\}$, times the FIG for that subdichotomy. Intuitively, for all initial beliefs and payoffs, the problem reduces at n large to the inference subproblem between the pair of states hardest to distinguish. A worst-case or 'minimax' inference rule thus arises endogenously as the asymptotically optimal one in a multi-state Bayesian problem.

Theorem 4 allows us to extend our theorems to K actions and M states. T81's Theorem 4.2 also leads to a minimax conclusion, but yields a ranking only for value levels. Unlike our many state extension, the proof in T81 critically depends on 0-1 losses.

Theorem 1 is valid with index $\rho_{\mathscr{E}} = \rho_{12}$ and constant $c_{MK} = (q_1+q_2)c_{2K}$. For by Section 5.A,

$$\begin{aligned} V^* - V(n) &= (q_1+q_2)[V_{12}^* - V_{12}(n)](1+O(\bar{\rho}_{\mathscr{E}}^n)) \\ &= (q_1+q_2)c_{2K}\frac{\rho_{12}^n}{\sqrt{n}}\left(1+O\left(\frac{1}{\sqrt{n}}\right)\right)(1+O(\bar{\rho}_{\mathscr{E}}^n)), \end{aligned}$$

which is $c_{MK}(\rho_{\mathscr{E}}^n/\sqrt{n})[1+O(1/\sqrt{n})]$, since $O(\bar{\rho}_{\mathscr{E}}^n)$ is vanishingly smaller than $O(1/\sqrt{n})$. As mentioned, Theorems 2 and 3 are likewise extended.

6. CONCLUSION

This paper offers contributions to classical questions about information. We work in the standard framework with a single opportunity to buy information, and have assumed that information units are cheap relative to payoffs, and can be purchased in large iid samples.

We have first ordered the value of information, by ranking almost all signals for their value in large samples. We then extended this ordering to the marginal value of information. Here the distinction between the statistical and economic approaches is most evident: The marginal value of information vanished as $1/n$ in the statistical deficiency sense (Mammen (1986)), but exponentially in our economic sense. We then derived a new logarithmic asymptotic formula describing the large demand for information, and provided an asymptotic error window of one unit—despite the unboundedness of demand. Finally, we have provided error bounds to underscore the validity of our results as approximations away from the limit. Our conclusions are simple and readily amenable to applications.

Department of Economics, Yale University, P.O. Box 208268, New Haven CT 06520-8268, U.S.A.; giuseppe.moscarini@yale.edu; www.econ.yale.edu/~mosca/mosca.html

and

Department of Economics, University of Michigan, 611 Tappan Street, Ann Arbor, MI 48109-1220, U.S.A.; lones@umich.edu; www.umich.edu/~lones

Manuscript received April, 2000; final revision received January, 2002.

APPENDIX: DEFERRED PROOFS

A. PROOF OF LEMMA 1

For ease in comparison with Section 5 of Bahadur and Rao (1960) (henceforth BR60), our proof is for the flip $\geq$ inequalities and for the opposite case of a process $S_n = \sum_{i=1}^n Y_i$ drifting to $-\infty$.

Lemma 2 in BR60 asserts that $\Pr(S_n \geq 0) = \rho^n I_n$, where $\rho \equiv E_Y[e^{\tau Y}]$ and $I_n \equiv \gamma\sqrt{n}\int_0^\infty e^{-\gamma\sqrt{n}x}[\Phi_n(x) - \Phi_n(0)]\,dx$ with $\gamma \equiv \sigma|\tau|$ and $\Phi_n(x) \equiv \Pr(S_n/\sigma\sqrt{n} < x)$. More strongly, we have $\Pr(S_n \geq \xi(q,u)) = \rho^n\gamma\sqrt{n}\int_{\xi_n}^\infty e^{-\gamma\sqrt{n}x}[\Phi_n(x) - \Phi_n(\xi_n)]\,dx$, with $\xi_n \equiv \xi(q,u)/\sigma\sqrt{n}$. The cdf's $\langle\Phi_n\rangle$ converge to the Gaussian cdf Φ, by the CLT. Under our assumptions, Y has a finite fourth moment. Hence by equation XVI.4.15 in Feller (1968) (with $r = 4$), we may write the *Edgeworth expansion* $\Phi_n(x) - \Phi(x) - \eta_n(x) = o(n^{-1})$, where

$$\eta_n(x) \equiv \Phi'(x)\left[\frac{R_3(x)}{\sqrt{n}} + \frac{R_4(x)}{n}\right]. \tag{7}$$

Here, $R_3(x), R_4(x)$ are polynomials depending only on the first four moments of Y. So $\eta_n'(x) = \Phi'(x)[\sqrt{n}P_3(x) + P_4(x)]/n$ for linear combinations P_3, P_4 of Hermite polynomials. For future reference, notice that $\bar{\eta}_n' \equiv \sup_{x\in\Re}|\eta_n'(x)| = O(1/\sqrt{n})$ *uniformly* in x, because the Gaussian density $\Phi'(x)$ swamps the polynomials, and since (7) holds uniformly in x.

Let $W_n|f| \equiv \gamma\sqrt{n}\int_{\xi_n}^\infty e^{-\gamma\sqrt{n}x}[f(x) - f(0)]\,dx$, so that $\rho^{-n}\Pr(S_n \geq \xi(q,u)) = W_n[\Phi_n] = W_n[\Phi] + W_n[\eta_n + o(1/n)]$ if the integrals finitely exist. Changing variables $y = \gamma\sqrt{n} + x$ yields:

$$\begin{aligned}
W_n[\phi] &= -e^{-\gamma\sqrt{n}x}[\Phi(x) - \Phi(\xi_n)]|_{\xi_n}^\infty + \int_{\xi_n}^\infty e^{-\gamma\sqrt{n}x}\frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}\,dx \\
&= 0 + e^{\frac{\gamma^2 n}{2}}\int_{\gamma\sqrt{n}+\xi_n}^\infty \frac{e^{-\frac{y^2}{2}}}{\sqrt{2\pi}}\,dy \\
&= e^{\frac{\gamma^2 n}{2}}[1 - \Phi(\gamma\sqrt{n} + \xi_n)].
\end{aligned}$$

Using the error function expansion $1 - \Phi(y) = (2\pi)^{-1/2}e^{-y^2/2}y^{-1}\{1 + O(y^{-2})\}$, valid for y large (see (54) in BR60), and $\gamma\xi(q,u)/\sigma = |\tau|\xi(q,u)$ by definition of γ, we then have

$$W_n[\phi] = \frac{e^{\frac{\gamma^2 n}{2}}e^{-\frac{(\gamma\sqrt{n}+\xi_n)^2}{2}}}{\sqrt{2\pi}}\frac{1}{\gamma\sqrt{n} + \xi_n}\left(1 + O\left(\frac{1}{n}\right)\right) = \frac{e^{-|\tau|\xi(q,u)}}{\gamma\sqrt{2\pi n}}\left(1 + O\left(\frac{1}{n}\right)\right).$$

For the other term $W_n[\eta_n + o(1/n)]$, fix $\varepsilon > 0$ and choose n large enough that $|o(1/n)]| < \varepsilon/n$ for all x, including ξ_n. Therefore, $o(1/n) - o(1/n) < 2\varepsilon/n$, and so

$$\begin{aligned}
W_n[\eta_n + o(1/n)] &< \gamma\sqrt{n}\int_{\xi_n}^\infty e^{-\gamma\sqrt{n}x}[\eta_n(x) - \eta_n(\xi_n)]\,dx + \gamma\sqrt{n}\int_{\xi_n}^\infty e^{-\gamma\sqrt{n}x}2\frac{\varepsilon}{n}\,dx \\
&= -e^{-\gamma\sqrt{n}x}[\eta_n(x) - \eta_n(\xi_n)]|_{\xi_n}^\infty + \int_{\xi_n}^\infty e^{-\gamma\sqrt{n}x}\eta_n'(x)\,dx + 2\frac{\varepsilon}{n}\left(e^{-\gamma\sqrt{n}x}|_{\xi_n}^\infty\right) \\
&< 0 + \bar{\eta}_n'\frac{e^{-\gamma\frac{\xi(q,u)}{\sigma}}}{\gamma\sqrt{n}} + 2\frac{\varepsilon}{n}e^{-\gamma\frac{\xi(q,u)}{\sigma}}
\end{aligned}$$

which is $O(1/n)$, since $\bar{\eta}_n' = O(1/\sqrt{n})$. Hence, multiplying by ρ^n, we find as claimed:

$$\begin{aligned}
\Pr(S_n \geq \xi(q,u)) &= \rho^n W_n[\Phi] + \rho^n W_n[\eta_n(x) + o(1/n)] \\
&= \frac{e^{-|\tau|\xi(q,u)}}{\gamma\sqrt{2\pi n}}\rho^n\left(1 + O\left(\frac{1}{n}\right)\right) + \rho^n O\left(\frac{1}{n}\right) \\
&= e^{-|\tau|\xi(q,u)}\frac{\rho^n}{\gamma\sqrt{2\pi n}}\left(1 + O\left(\frac{1}{\sqrt{n}}\right)\right).
\end{aligned}$$

Q.E.D.

B. Omitted Parts of Proof of Theorem 3

Claim 1: *If* $D(p) = \arg\max_{n\in\mathbb{N}} V(n) - np$, *and* $\underline{n}(p) = \inf D(p)$, *then* $\lim_{p\to 0} \underline{n}(p) = \infty$.

Proof: First note that the demand correspondence is decreasing at *all* n. Indeed, $V(n)$ is monotone in its ordered argument n (marginal value is positive), and so is quasisupermodular. Hence, $V(n) - np$ satisfies a single-crossing property (SCP) in $(n, -p)$, and thus the set $D(p)$ of maximizers $n(p)$ is nondecreasing in $-p$, i.e nonincreasing in p.[13] In particular, $\inf D(p)$ and $\sup D(p)$ are monotonic. So the infimum $\underline{n}(p)$ has a limit, say B. If $B < \infty$, then $\underline{n}(p) \leq B < \infty$ for all p. Since $V(n)$ is strictly increasing in n, we may choose $\bar{n}$, such that $\varepsilon \equiv V(\bar{n}) - V(B) > 0$. Since $\underline{n}(p) + \bar{n} \geq \bar{n}$ and $\underline{n}(p) \leq B$ for all p, $V(\underline{n}(p) + \bar{n}) - V(\underline{n}(p)) \geq \varepsilon > 0$ for all p. As $\underline{n}(p)$ is a maximizer, $V(\underline{n}(p)) - \underline{n}(p)p \geq V(\underline{n}(p) + \bar{n}) - (\underline{n}(p) + \bar{n})p$. Altogether, $\bar{n}p \geq V(\underline{n}(p) + \bar{n}) - V(\underline{n}(p)) \geq \varepsilon$ for all p small enough, contradicting $\bar{n} < \infty$. *Q.E.D.*

Claim 2: *For almost all* p, $D(p)$ *is single-valued*.

Proof: Maximizing a function $V(n) - np$ over integers yields multiple solutions n', n'' only if $V(n') - n'p = V(n'') - n''p$. For each pair n', n'', a unique such p is implied. This set is clearly countable. *Q.E.D.*

C. Proof of Theorem 4

Recall that we have M states and K actions, with action a_i strictly best in state θ_i. Let $h_i = \arg\max_{\ell\neq i} \rho_{i\ell}$, so that θ_{h_i} is the state hardest to disentangle asymptotically from θ_i. Then $h_{h_i} = i$ by symmetry of the Hellinger transform, while $\rho_{\mathscr{E}} \equiv \rho_{12}$ is the unique maximand across $\rho_{i\ell}$. Also, for $h = 1, 2, \ldots, M$, let $Z_{hi}(n) \equiv f_h(X^n)/f_i(X^n) \in (0, \infty)$ denote the likelihood ratio (LR) of the n observations X^n in state θ_h versus state $\theta_i \neq \theta_h$.

Claim 3: *For every* $\varepsilon > 0$ *and* $(b_1, \ldots, b_M) \in \mathbb{R}^M$: $\Pr(\sum_{h\neq i} b_h Z_{hi}(n) \geq \varepsilon|\theta_i) = O(\rho^n_{ih_i}/\sqrt{n})$.

Proof: If all $b_h \leq 0$ for all $h \neq i$, we're done, since $Z_{hi}(n)$ is positive (being a LR), so the chance in the claim is zero. Assume some $b_h > 0$. If $\sum_{h\neq i} b_h Z_{hi}(n) > \varepsilon$, at least one summand exceeds $\varepsilon/(M-1)$. Since $\varepsilon - (M-1)\sum_{j\neq i, b_j\leq 0}(b_j/b_h)Z_{ji}(n) > \varepsilon$, we have

$$\Pr\left(\sum_{h\neq i} b_h Z_{hi}(n) > \varepsilon|\theta_i\right) \leq \sum_{h\neq i, b_h>0} \Pr\left(Z_{hi}(n) \geq \frac{\varepsilon}{(M-1)b_h} - \sum_{j\neq i, b_j\leq 0} \frac{b_j}{b_h} Z_{ji}(n)|\theta_i\right).$$

Now, each LR sequence $Z_{hi}(n) \to 0$ a.s. in state θ_i, and the chance it exceeds a positive threshold vanishes at rate $\rho^n_{ih}/\sqrt{n}$. Thus, the sum is $\sum_{h\neq i} O(\rho^n_{ih}/\sqrt{n}) = O(\rho^n_{ih_i}/\sqrt{n})$, by Lemma 1, and because the rate is (weakly) slowest for $h = h_i$, by definition of h_i. *Q.E.D.*

We next show that the chance of taking any action $a_j \neq a_i$ in state θ_i vanishes no slower than $\rho^n_{ih_i}/\sqrt{n}$. So when $i = 1, 2$, this chance is eventually the same as when only states θ_1 and θ_2 exist, that subinference problem being the most resilient to Large Numbers. To this end, for $i = 1, 2$, let $\rho_{ik_i} = \max_{h\neq 1,2} \rho_{ih}$ denote the largest pairwise index between state θ_i and states $\theta_h, h > 2$, by assumption strictly less than $\rho_{\mathscr{E}} \equiv \rho_{21} = \max_{h\neq i} \rho_{ih}$. That is, $\max\langle \rho_{1k_1}, \rho_{2k_1}, \max_{\ell>2} \rho_{\ell h_\ell}\rangle < \rho_{\mathscr{E}}$. Next, let α_{ji} denote the chance of taking action a_j in state $\theta_i \neq \theta_j$, for $j = 1, 2, \ldots, K$—hence a chance of error in the larger (M, K) problem. Finally, consider a subdecision problem—a *double subdichotomy*—with the prior focused only on states θ_k and θ_i, with chances $(q_k, q_i)/(q_i + q_k)$, and where only actions a_j and a_i are admissible. In this problem, let $\alpha^{(ki)}_{ji}$ be the chance that action a_j yields higher expected utility than action a_i.

[13] See Theorem 2.8.6 in Topkis (1998). We thank a referee for drawing our attention to this implication.

CLAIM 4: $\alpha_{ji} = O(\rho^n_{ih_i}/\sqrt{n})$, *and so* $\alpha_{ji} \leq \alpha^{(12)}_{ji}(1 + O(\rho^n_{ik_i}/\rho^n_{\mathscr{E}}))$ *for* $i = 1, 2$.

PROOF: Denote by $E_{jk} = E_{jk}(X^n)$ the event that action a_j yields a higher expected utility than a_k after observing X^n in the M-state, K-action problem. That is,

$$E_{jk} = \left\{ \sum_{h=1}^{M} \frac{q_h f_h(X^n) u_{jh}}{\sum_{l=1}^{M} q_l f_l(X^n)} \geq \sum_{h=1}^{M} \frac{q_h f_h(X^n) u_{kh}}{\sum_{l=1}^{M} q_l f_l(X^n)} \right\} = \left\{ \sum_{h=1}^{M} q_h f_h(X^n) u_{jh} \geq \sum_{h=1}^{M} q_h f_h(X^n) u_{kh} \right\}.$$

So the chance α_{ji} of taking action a_j in state θ_i cannot exceed the chance that action a_j beats a_i, the best action in that state: namely, $\alpha_{ji} = \Pr(\bigcap_{k \neq j} E_{jk} | \theta_i) \leq \Pr(E_{ji}|\theta_i)$. Then

$$E_{ji} = \left\{ q_i(u_{ji} - u_{ii}) + q_{h_i} Z_{h_i i}(n)\big(u_{jh_i} - u_{ih_i}\big) + \sum_{h \neq i, h_i} q_h Z_{hi}(n)(u_{jh} - u_{ih}) \geq 0 \right\}$$

$$\subseteq \big\{ q_{h_i} Z_{h_i i}(n)\big(u_{jh_i} - u_{ih_i}\big) > -q_i(u_{ji} - u_{ii}) - \varepsilon \big\} \cup \left\{ \sum_{h \neq i, h_i} q_h Z_{hi}(n)(u_{jh} - u_{ih}) \geq \varepsilon \right\},$$

which we write as $E_{ji} \subseteq F^{\varepsilon}_{ji} \cup G^{\varepsilon}_{ji}$. Now, $\Pr(F^0_{ji}|\theta_i)$ is the chance $\alpha^{(h_i i)}_{ji}$ that action a_j beats a_i in the double subdichotomy of state θ_{h_i} vs. θ_i, and so $\Pr(F^0_{ji}|\theta_i) = O(\rho^n_{ih_i}/\sqrt{n})$, by the double dichotomy theory. Next, if $b_h \equiv q_h(u_{jh} - u_{ih})$ and $\varepsilon > 0$, Claim 3 yields $\Pr(\sum_{h \neq i, h_i} b_h Z_{hi}(n) > \varepsilon | \theta_i) = O(\rho^n_{ik_i}/\sqrt{n})$. Since $\rho_{ih_i} = \max_{h \neq i} \rho_{ih} \geq \max_{h \neq i, h_i} \rho_{ih} \equiv \rho_{ik_i}$, we have $\Pr(G^{\varepsilon}_{ji}|\theta_i) = O(\rho^n_{ih_i}/\sqrt{n})$. Finally, $\alpha_{ji} \leq \Pr(F^0_{ji}|\theta_i) + \Pr(G^{\varepsilon}_{ji}|\theta_i) = O(\rho^n_{ih_i}/\sqrt{n})$.

Next, assume $i = 1$. Since $E_{j1} \subseteq F^0_{j1} \cup G^{\varepsilon}_{j1}$, we have $\alpha_{j1} \leq \Pr(F^0_{j1}|\theta_1) + \Pr(G^{\varepsilon}_{j1}|\theta_1) = \alpha^{(21)}_{j1} + \Pr(G^{\varepsilon}_{j1}|\theta_1)$, since $h_1 = 2$. Applying Claim 3 to $h > 2$, and using $\rho_{1k_1} = \max_{h>2} \rho_{1h}$:

$$\alpha_{j1} \leq \alpha^{(21)}_{j1} + \Pr\left(\sum_{h>2} b_h Z_{h1}(n) \geq \varepsilon | \theta_1 \right) = \alpha^{(21)}_{j1}\big(1 + O\big(\rho^n_{1k_1}/\rho^n_{\mathscr{E}}\big)\big)$$

where the last equality owes to $\alpha^{(21)}_{j1} = O(\rho^n_{\mathscr{E}}/\sqrt{n})$, true by the dichotomy theory. *Q.E.D.*

Finally, we complete the proof of the equality in Theorem 4, namely that $V^* - V(n) = (q_1 + q_2)[V^*_{12} - V_{12}(n)](1 + o(\bar{\rho}^n_{\mathscr{E}}))$ by establishing two separate inequalities.

CLAIM 5: $(q_1 + q_2)[V^*_{12} - V_{12}(n)] \leq V^* - V(n)$.

PROOF: We consider an auxiliary signal with $M - 1$ outcomes $\{x_{12}, x_3, x_4, \ldots, x_M\}$, with the following likelihoods. $\Pr(x_{12}|\theta_i) = 1$ if $i = 1$ or $i = 2$ and zero otherwise; and for $i > 2$, $\Pr(x_i|\theta_j) = 1$ if $i = j$ and 0 otherwise. This signal is perfectly informative of the true state, except in states θ_1 and θ_2: After the realization x_{12}, the $\mathscr{DM}$ knows $\theta \in \{\theta_1, \theta_2\}$, with chances $q_1/(q_1 + q_2)$ and $q_2/(q_1 + q_2)$; after any realization x_j, for $j > 2$, the true state must be θ_j. In the latter case, the $\mathscr{DM}$ takes action a_j and earns u_{jj}. Let the $\mathscr{DM}$ observe this signal first and then observe n independent draws of the original experiment. The ex ante unconditional chance of realizing x_{12} is $\sum_i q_i \Pr(x_{12}|\theta_i) = q_1 \Pr(x_{12}|\theta_1) + q_2 \Pr(x_{12}|\theta_2) + \sum_{i>2} q_i 0 = q_1 + q_2$. Let $V_{12}(n)$ denote the value of running the 12 dichotomy n times with prior $q_1/(q_1 + q_2)$. Observing the informative auxiliary signal cannot hurt the $\mathscr{DM}$, so that $V(n) \leq (q_1 + q_2)V_{12}(n) + \sum_{j>2} q_j u_{jj}$. The full information payoff of the 12 dichotomy is $V^*_{12} = (q_1 u_{11} + q_2 u_{22})/(q_1 + q_2)$, and so $(q_1 + q_2)V^*_{12} + \sum_{j>2} q_j u_{jj} = \sum q_j u_{jj} = V^*$. Altogether:

$$V^* - V(n) \geq \left[(q_1 + q_2)V^*_{12} + \sum_{j>2} q_j u_{jj} \right] - \left[(q_1 + q_2)V_{12}(n) + \sum_{j>2} q_j u_{jj} \right]. \qquad \textit{Q.E.D.}$$

Define $\bar{\rho}_{\mathscr{E}} \equiv \max\langle \rho_{1k_1}, \rho_{2k_1}, \max_{\ell>2} \rho_{\ell h_\ell} \rangle / \rho_{\mathscr{E}}$. As noted earlier, $\bar{\rho}_{\mathscr{E}} < 1$.

CLAIM 6: $V^* - V(n) \le (q_1+q_2)[V_{12}^* - V_{12}(n)](1+O(\bar{\rho}_{\mathscr{E}}^n))$.

PROOF: By definition, $V^* - V(n) = \sum_{i=1}^M q_i \sum_{j=1}^K \alpha_{ji}(u_{ii} - u_{ji})$. Hence by Claim 4:

$$\begin{aligned} V^* - V(n) &\le \sum_{i=1}^{2} q_i \sum_{j=1}^{K} (u_{ii} - u_{ji}) \alpha_{ji}^{(12)} (1 + O(\bar{\rho}_{\mathscr{E}}^n)) + \sum_{\ell=3}^{M} q_\ell O\big(\rho_{\ell h_\ell}^n / \sqrt{n}\big) \\ &= (q_1+q_2) \sum_{i=1}^{2} \frac{q_i}{q_1+q_2} \sum_{j=1}^{K} (u_{ii} - u_{ji}) \alpha_{ji}^{(12)} \big(1 + O(\bar{\rho}_{\mathscr{E}}^n)\big) \\ &= (q_1+q_2)[V_{12}^* - V_{12}(n)]\big(1 + O(\bar{\rho}_{\mathscr{E}}^n)\big) \end{aligned}$$

for since $\bar{\rho}_{\mathscr{E}} < 1$, the $O(\rho_{\ell h_l}^n / \sqrt{n})$ terms are of smaller order than the lead $\alpha_{ji}^{(12)}$ terms. *Q.E.D.*

REFERENCES

ATHEY, S., AND J. LEVIN (2001): "The Value of Information in Monotone Decision Problems," Mimeo, Stanford.

BAHADUR, R. R., AND R. R. RAO (1960): "On Deviations of the Sample Mean," *Annals of Mathematics and Statistics*, 31, 1015–1027.

BLACKWELL, D. (1953): "Equivalent Comparison of Experiments," *Annals of Mathematics and Statistics*, 24, 265–272.

CHADE, H., AND E. SCHLEE (2001): "Another Look at the Radner-Stiglitz Non-concavity in the Value of Information," forthcoming in *Journal of Economic Theory*.

CHERNOFF, H. (1952): "A Measure of Asymptotic Efficiency for Tests of a Hypothesis Based on the Sum of Observations," *Annals of Mathematical Statistics*, 23, 493–507.

CRAMÉR, H. (1938): "Sur un Nouveau Théorème-limite de la Théorie des Probabilités," *Actualités Scientifiques et Industrielles*, No. 736. Paris: Hermann.

FELLER (1968): *An Introduction to Probability Theory and Its Applications*, Vol. I. New York: John Wiley & Sons, Inc., Third Edn.

LE CAM, L. (1964): "Sufficiency and Approximate Sufficiency," *Annals of Mathematical Statistics*, 35, 1419–1455.

LEHMANN, E. L. (1988): "Comparing Location Experiments," *Annals of Statistics*, 16, 521–533.

MAMMEN, E. (1986): "The Statistical Information Contained in Additional Observations," *Annals of Statistics*, 14, 665–678.

MOSCARINI, G., AND L. SMITH (2001): "The Optimal Level of Experimentation," *Econometrica*, 69, 1629–1644.

PERSICO, N. (2000): "Information Acquisition in Auctions," *Econometrica*, 68, 135–148.

RADNER, R., AND J. STIGLITZ (1984): "A Nonconcavity in the Value of Information," in *Bayesian Models in Economic Theory*, ed. by M. Boyer and R. Kihlstrom. New York: Elsevier Science Publishers, pp. 33–52.

SANDRONI, A. (2000): "Do Markets Favor Agents Able to Make Accurate Predictions?" *Econometrica*, 68, 1303–1341.

SCHERVISCH, M. (1995): *Theory of Statistics*. New York: Springer-Verlag.

SHAPIRO, C., AND H. R. VARIAN (1999): *Information Rules: A Strategic Guide to the Network Economy*. Boston, MA: Harvard Business School Press.

TOPKIS, D. (1998): *Supermodularity and Complementarity*. Princeton, N.J.: Princeton University Press.

TORGERSEN, E. (1981): "Measures of Information Based on Comparison with Total Information and with Total Ignorance," *Annals of Statistics*, 9, 638–657.

——— (1991): *Comparison of Statistical Experiments*, Vol. 36 of *Encyclopedia of Mathematics and its Applications*. Cambridge: Cambridge University Press.