Available online at www.sciencedirect.com
ScienceDirect
Journal of Economic Theory 138 (2008) 21–50
JOURNAL OF Economic Theory
www.elsevier.com/locate/jet

The demand for information: More heat than light

Jussi Keppo[a], Giuseppe Moscarini[b], Lones Smith[c,*]

[a]Industrial and Operations Engineering Department, University of Michigan, 611 Tappan St., Ann Arbor, MI 48109-1220, USA
[b]Economics Department, Yale University, 28 Hillhouse Ave., New Haven, CT 06510 Cowles Foundation and NBER
[c]Economics Department, University of Michigan, 611 Tappan St., Ann Arbor, MI 48109-1220, USA

Received 25 July 2005; final version received 21 March 2007
Available online 23 May 2007

Abstract

This paper produces a theory of value for Gaussian information with two states and two actions, tracing the solution of the option pricing formula, but for the process of beliefs. We derive explicit formulas for the value of information. The marginal value is convex and rising, concave and peaking, and finally convex and falling. As the price falls, demand is initially zero, and eventually logarithmic. Its elasticity exceeds one, and falls to zero with the price. Demand is hill-shaped in beliefs, zero at extremes. Our results approximate models where information means the sample size for weak discrete informative signals.

JEL classification: D81; G13

Keywords: Heat equation; Information; Garbling; Demand curve; Elasticity; Martingale; Non-concavity; Option pricing; Diffusion process; Ito's Lemma

1. Introduction

Information acquisition is an irreversible process. One cannot recapture the pristine state of ignorance once apprised of any given fact. Heat dissipation also obeys the arrow of time: the heat equation in physics describing its transition is not time symmetric. This paper argues that this link is not merely philosophical. In static models of Bayesian information acquisition, the value function of prior beliefs and information quantity obeys an inhomogeneous heat equation. We show that a nonlinear transformation of the value function and beliefs exactly obeys the heat equation. This paper exploits this fact and crafts a global theory of information and demand. For a binary state

* Corresponding author. Fax: +1 734 764 2769.
E-mail address: lones@umich.edu (L. Smith).

doi:10.1016/j.jet.2007.03.003

world, we derive explicit formulas that provide the bigger picture on the famous nonconcavity of information, and the unique demand curve that it induces: we characterize the "choke-off demand" level, and also make many novel findings—e.g., demand elasticity is monotonically falling to zero.

Information is an important good for decision-making, and therefore in game theory and general equilibrium analysis. Yet information is also a poorly understood good. This first of all owes to a lack of agreed units. Blackwell's Theorem only considers one signal, for instance. We thus start by measuring information in units corresponding to signal sample sizes, or equivalently, the precision of a normally distributed signal. This is the foundation of our entire theory.

Consider a static binary action decision problem where the decision cannot be postponed, there is no option value of waiting, one must act now and bear the consequences. For instance, a pregnant woman decides on an amniocentesis to detect fetal genetic defects, or a company decides whether to enter a market before its competitors. More formally, we have the standard $\vee$-shaped payoff function of the belief in the high state: a decision maker takes the low action left of a cross-over belief, and the high action right of it. We show that the ex post value of information is a multiple of the payoff to a call option whose strike price is the cross-over belief. Our analysis not surprisingly traces the development of the option pricing formula.

One technical contribution opens the door to this option story, showing that a belief process behaves like the price process of the underlying option. Yet the belief process is less tractable than the geometric Brownian motion for asset prices, and our transformation thus more complex. We produce in Lemmas 1–4 a transformation jointly of time and beliefs yielding a detrended log likelihood ratio process. This is the unique process sharing two critical characteristics: first, it has unit diffusion coefficient (variance), and second it maintains a one-dimensional stochastic analysis. This yields in Lemma 5 a simple transition law for beliefs. This density is conveniently proportional to its time derivatives (Lemma 6).

Using the belief process, the payoff function satisfies an inhomogeneous heat equation (Lemma 7). This equation reveals that information has positive marginal value iff payoffs are convex in beliefs. In our critical innovation, we perform a change of variables, blending time and log-likelihood ratios, and a nonlinear transformation of payoffs to produce the standard heat equation. This leads us to the option pricing exercise: the heat equation approach of [1] and the martingale measure tack of [4] in Lemma 8. We juxtapose both proofs for instructive value. Either way, Theorem 1 explicitly expresses the value of information in terms of the normal distribution.

We then turn to our substantive findings. Theorem 2 expresses the marginal value of information in terms of the derivatives of the transition belief density. This reduces the analysis of the value derivative to a polynomial in the reciprocal demand. Using this, Corollary 1 finds that the marginal value is initially zero—the nonconcavity—as found in [12], and rigorously formalized in [2]. The sufficient conditions in [2] for this result, the sharpest known to date, do not encompass our model. So this is a novel result: in a finite decision problem, the value of gaussian information, as measured by its precision, is initially convex. By Theorem 3, the marginal value convexly rises, is then concave and hill-shaped, and finally convex and falling—as in Fig. 1.

So with linear prices, information demand vanishes at high prices, before jumping to a positive level (Theorem 4) strictly below the peak marginal value. The Law of Demand then kicks in, and demand coincides with the marginal value schedule. One never buys just a little information. Theorem 5 quantifies the nonconcavity: the minimal expenditure on average is about 2.5% of the expected payoff stakes.

In Theorem 6, we find that information demand is hill-shaped in beliefs, unlike in the dynamic model with impatience of [10]. Also, it jumps down to 0 near beliefs 0 and 1, when the choke-off

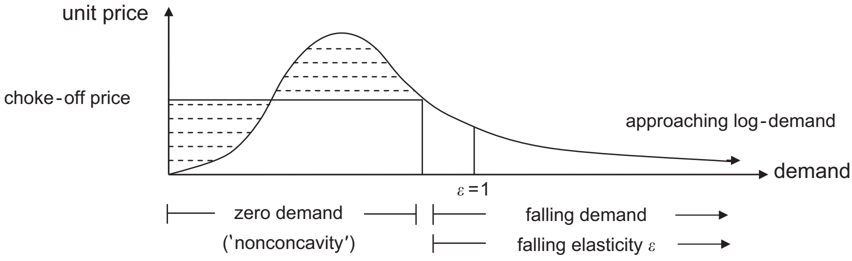

Fig. 1. Marginal value of information.

demand binds. Here we see an interesting contrast with sequential learning models, because our thresholds owe to global optimality. Completing this picture, Theorem 7 finds that information demand is hill-shaped in beliefs (quasi-concave, precisely)—opposite to [10].

A novel topic we explore is the demand elasticity. Theorem 8 asserts that the information demand is initially elastic at interior beliefs; the elasticity is globally falling in price, and is vanishing in the price. This dovetails with what comes next. We revisit in Theorem 9 the large demand analysis of [9] now quickly via our formulas rather than large deviation theory. [9] also measure information by sample size, but assume cheap discrete information, and not our continuous signals. The marginal value of information eventually vanishes exponentially fast, producing the logarithmic demand of [9] at small prices. We sharpen the demand approximation, and find that it is monotonically approached.

Our Gaussian information is generated by the time that one observes a Brownian motion with state-dependent drift, but state-independent diffusion coefficient. However, consider a discrete model where the decision-maker chooses how many conditional iid signals to draw, with a state-independent variance. Theorem 10 shows that Bayes' updating weakly approaches the continuous time Brownian filter as the signal strength vanishes, and the implied likelihood ratio converges to 1. We prove that garbled signals have precisely this property. We also show in Theorem 11 that the demand formulas and value of information approximate the discrete formulas. In summary, the paper produces an approximate value of information and demand curve for all models where information is measured by the size of a sample of discrete information "bits'', namely weak signals, with (asymptotically) known variance.

The experimentation literature aside, we know of one related information demand paper. In [7], a consumer faces a linear price for precision of a Gaussian signal given a (conjugate) Gaussian prior. For a hyperbolic utility function, he can write utility as a function of signals and avoid computing the density of posteriors. Our theory of the binary state model is not driven by such linearity.

We next lay out the model, and develop the results on beliefs, the value and marginal value of information, demand, and weak discrete signals.

2. The model

2.1. The decision problem

Assume a one-shot decision model, where a decision maker (*DM*) chooses how much information to buy, and then acts. For simplicity, we assume two actions A, B, whose payoffs π_A^θ, π_B^θ

depend on the payoff-relevant state of the world $\theta = L, H$. Action B is best in state H and action A is best in state L: $0 \leqslant \pi_A^H < \pi_B^H$ and $\pi_A^L \geqslant \pi_B^L \geqslant 0$.[1] Since the *DM* has the prior belief $q \in (0, 1)$ that $\theta = H$, the convex $\vee$-shaped expected payoff function is

$$u(q) = \max\langle q\pi_A^H + (1-q)\pi_A^L, q\pi_B^H + (1-q)\pi_B^L\rangle \equiv \max\langle \pi_A^L + mq, \pi_B^L + Mq\rangle \tag{1}$$

thereby defining $m = \pi_A^H - \pi_A^L$ and $M = \pi_B^H - \pi_B^L$. We assume no dominated actions, so that payoffs have a kink at a *cross-over belief* $\hat{q} = (\pi_A^L - \pi_B^L)/(M - m) \in (0, 1)$.

The maximum *payoff stakes* here are $(\pi_B^H - \pi_A^H) + (\pi_A^L - \pi_B^L) = M - m > 0$. The *DM* never incurs a payoff loss greater than $M - m$ from a wrong action choice; this bound is tight when either difference $\pi_B^H - \pi_A^H \geqslant 0$ or $\pi_A^L - \pi_B^L \geqslant 0$ vanishes.

2.2. *The standard information acquisition model*

Given is a probability space $(\Omega, \mathcal{F}, P)$, where Ω is a set, $\mathcal{F}$ a σ-algebra of subsets of Ω, and P a probability measure on $\mathcal{F}$. This space captures all uncertainty, including the state $\theta = L, H$.

Before deciding, the *DM* can obtain any level $t \geqslant 0$ of information about the state θ. While more information could plausibly connote better quality information, we mean that the *DM* with information level t_2 knows strictly more about the state of the world than does the *DM* with information level $t_1 \leqslant t_2$. So assume a filtration $\{\mathcal{F}_t : t \in [0, \infty)\}$, so that the σ-algebras are nested $\mathcal{F}_{t_1} \subset \mathcal{F}_{t_2} \subseteq \mathcal{F}$ when $t_1 < t_2$. The *DM* observes $\mathcal{F}_t$, updates her beliefs to $q(t) = P(H|\mathcal{F}_t)$ from the prior $q(0) = q$.

The *ex ante expected payoff* (prior to seeing $\mathcal{F}_t$) is $u(t, q) = E[u(q(t))|q(0) = q]$, and the *value of information* is $v(t, q) = u(t, q) - u(q)$; namely, the expected increase in utility from observing $\mathcal{F}_t$. Faced with a constant marginal cost $c > 0$ of information, the *DM* can choose the observation 'time' t at cost ct. The net payoff given the level t is $v(t, q) - ct$. This is maximized by choosing the information level $\tau(c, q) > 0$, which is our demand schedule. Finally, the *DM* chooses the best action.

2.3. *The units of continuous information*

We actually have a filtration in mind. Let the *DM* observe the time-t realization of a Brownian motion $X(\cdot)$ with drift $\pm\mu$ in states H, L, respectively, and constant diffusion coefficient $\sigma > 0$. Thus, the signal is twice as informative when t doubles—just as is true for the sample size of conditionally iid signals. We show in Section 7 that this approximates discrete bit sampling models. Throughout this paper, time is used only metaphorically, but really means the level of information: the model is formally static!

By Theorem 9.1 in [8], when observing the realizations of the Brownian Motion $X(t)$ in continuous time, the belief $\tilde{q}(t) = P(H|\mathcal{F}_t)$ obeys the Bayes filter[2] $d\tilde{q}(t) = \zeta\tilde{q}(t)(1-\tilde{q}(t))\,dW(t)$, where $\zeta \equiv 2\mu/\sigma$ is the *signal/noise ratio*, and $W(\cdot)$ is a standard Wiener process w.r.t. the un-

[1] We assume without loss of generality for simplicity that $\pi_A^H = 0$, since the decision must be made. An analogous choice of $\pi_B^H = 1$ is not allowed later on, without also scaling the cost function.

[2] For an intuition, observe that by Bayes rule,

$$q(t+\Delta) - q(t) = \frac{q(t)P(\xi\ signal|R)}{q(t)P(\xi\ signal|R) + (1-q(t))P(\xi\ signal|L)} - q(t) \propto q(t)(1-q(t)).$$

conditional measure P. If we define $q(t) = \tilde{q}(t/\zeta^2)$, then[3] $(dq(t))^2 = (d\tilde{q}(t))^2/\zeta^2 = q(t)^2 (1-q(t))^2\, dt$, and thus

$$dq(t) = q(t)(1-q(t))\, dW(t). \tag{2}$$

We henceforth set $\zeta = 1$ *and compute the time* $\hat{t}$ *with any* $\hat{\zeta} > 0$ *from* $\hat{t} = t/\hat{\zeta}^2$.

3. Beliefs and log likelihood ratios

We begin by describing the limit behavior of beliefs $q(\cdot)$.

Lemma 1 (*Long run beliefs*). *The belief process in* (2) *satisfies*

(a) $P\left[\inf\{t \geqslant 0 \mid q(t) = 0 \text{ or } 1\} = \infty\right] = 1,$

(b) $P\left[\lim_{t\to\infty} q(t) = 0\right] = 1 - P\left[\lim_{t\to\infty} q(t) = 1\right] = 1 - q.$

So the probability that $q(t) \notin (0, 1)$ *in finite time is zero and* $q(\infty) = 0$ *or* 1.

The proof of (a) is in the appendix: $q(\cdot)$ avoids the boundary as the diffusion coefficient in (2) vanishes quickly near 0, 1. The martingale property gives part (b).

We seek to derive from posterior beliefs a tractable process that contains the same information. In particular, we aim for a simple standard Wiener process. While the logLR of state L to state H is a martingale on state H, this is not useful unconditionally. We instead *first* subtract the deterministic portion of the drift from the logLR, and *then* change measure from P to a conditional measure Q that yields a martingale. In Lemma 2, we find the unique monotone transformation of posterior beliefs with a unit diffusion coefficient. Next, in Lemma 4, we change probability measure so that this transformation retains the martingale property. There is a degree of freedom here, which we resolve in Lemma 3 on grounds of tractability.

Lemma 2 (*Likelihood ratios*). *Let* $z(t) = \lambda(t, q(t))$, *where* $\lambda \in C^2$. *If the diffusion coefficient of* $z(\cdot)$ *is one then* $\lambda(t, q(t)) = A(t) + \log\left(\frac{q(t)}{1-q(t)}\right)$, *where* $|A(t)| < \infty$.

Proof. Using Ito's Lemma we get

$$dz(t) = \left(\lambda_t + \tfrac{1}{2}\lambda_{qq} q(t)^2(1-q(t))^2\right) dt + \lambda_q q(t)(1-q(t))\, dW(t). \tag{3}$$

Solving $\lambda_q q(t)(1-q(t)) = 1$ yields $\lambda(t, q(t)) = A(t) + \log\left(\frac{q(t)}{1-q(t)}\right)$. □

This lemma is intuitive, since Bayes rule is multiplicative in likelihood ratios, and therefore additive in log-likelihood ratios. Substituting from (3), we then find

$$dz(t) = \left(A'(t) - \frac{1-2q(t)}{2}\right) dt + dW(t) \equiv v(t)\, dt + dW(t), \tag{4}$$

[3] To justify the first inequality, $E[W^2(t/\zeta^2)] = t/\zeta^2$ and thus $dW^2(t/\zeta^2) = dt/\zeta^2$.

where we have implicitly defined $v(t)$. Next, define the probability measure Q on $(\Omega, \mathcal{F}_t)$ by the Radon–Nikodym derivative:

$$\frac{dQ}{dP} = R(t) = \exp\left(-\frac{1}{2}\int_0^t v^2(s)\,ds - \int_0^t v(s)\,dW(s)\right). \tag{5}$$

Lemma 3 (*Radon–Nikodym derivative*). $R(t) = q(t)/q$ *iff* $A(t) = -t/2$.

This result is important as it does not introduce a new stochastic process. Any different R–N derivative would yield two imperfectly correlated stochastic processes $q(\cdot)$ and $R(\cdot)$, and render derivation of our results exceedingly difficult. A unique change of measure maintains the unidimensionality of the stochastic process.

To prove Lemma 3, write $Y(t) = -\frac{1}{2}\int_0^t (1-q(s))^2\,ds + \int_0^t (1-q(s))\,dW(s)$. Guessing $q(t) = qe^{Y(t)}$ correctly yields, using Ito's Lemma, $dq(t) = q(t)\,dY(t) + \frac{1}{2}q(t)(dY(t))^2 = q(t)(1-q(t))\,dW(t)$—namely, our belief filter (2). In other words,

$$q(t) = q\exp\left(-\frac{1}{2}\int_0^t (1-q(s))^2\,ds + \int_0^t (1-q(s))\,dW(s)\right).$$

So motivated, if we set $v(s) = q(s) - 1$ in (5), then we get the result of Lemma 3.

Eq. (4) implies that $A(t) = -t/2$, which we henceforth assume. Thus,

$$z(t) = \lambda(t, q(t)) = \log\left(\frac{q(t)}{1-q(t)}\right) - \frac{1}{2}t \quad \Leftrightarrow \quad q(t) = \ell(t, z(t)) = \frac{1}{e^{-\frac{1}{2}t - z(t)} + 1}. \tag{6}$$

We next exploit the fact that the partially de-trended log-likelihood ratio $z(\cdot)$ is a Q-Wiener process—by Girsanov's theorem (see [11, pp. 155–156]).

Lemma 4 (*Log likelihood ratio*). *The process* $z(\cdot)$ *obeys* $dz(t) = d\hat{W}(t)$, *where the process* $\hat{W}(t) = \int_0^t (q(s)-1)\,ds + W(t)$ *is Wiener under the probability measure* Q. *So* $z(\cdot)$ *is a* Q-*martingale, and the pdf for transitions* $z \mapsto y$ *in time* $t > 0$ *equals*:

$$\frac{1}{\sqrt{t}}\phi\left(\frac{y-z}{\sqrt{t}}\right) = \frac{1}{\sqrt{2\pi t}}e^{\frac{-(y-z)^2}{2t}} \tag{7}$$

for all $(t, z, y) \in (0, \infty) \times \mathbb{R}^2$, *where* $\phi(y) = \frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}$ *is the standard normal pdf.*

We now derive the belief transition pdf

$$\xi(t, q, r) = \frac{\partial}{\partial r}P(q(t) \leqslant r | q(0) = q),$$

using the transition pdf (7) for transformed log-likelihood ratios $z(t)$ (Fig. 2). We then repeatedly exploit this belief transition pdf to derive the formula for the value and marginal value of information, as well as the properties of the demand function. We next produce an extremely useful exact formula for it.

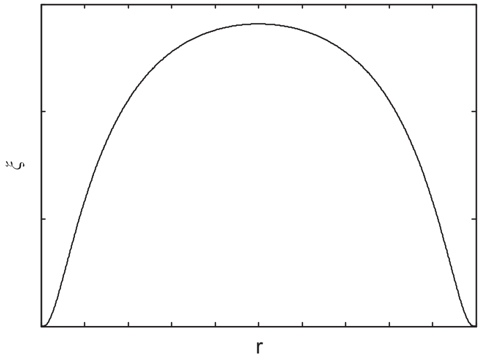

Fig. 2. Transition probability function. We plot the symmetric pdf $\xi(t, q, r)$ for transitions from $q = 0.5$ to any belief r after an elapse time $t = 1$.

Lemma 5 (*Beliefs*). *The transition pdf of beliefs $q(t)$ is given by*

$$\xi(t, q, r) = \frac{q\phi\left(-\frac{1}{2}\sqrt{t} + \frac{1}{\sqrt{t}}L(r, q)\right)}{r^2(1-r)\sqrt{t}} = \sqrt{\frac{q(1-q)}{r^3(1-r)^3 2\pi t}}\, e^{-\frac{1}{8}t - \frac{1}{2t}L^2(r,q)} \tag{8}$$

for all $(t, q, r) \in (0, \infty) \times (0, 1) \times (0, 1)$, *where* $L(r, q) = \log\left(\frac{r(1-q)}{(1-r)q}\right)$.

Proof. Fix a measurable real function $\eta(\cdot)$. Then $\int_0^1 \xi(t, q, r)\eta(r)\,dr$ equals

$$E_q[\eta(q(t))] = qE_q\left[\frac{q(t)}{q}\frac{\eta(q(t))}{q(t)}\right] = qE_q^Q\left[\frac{\eta(q(t))}{q(t)}\right] = qE^Q_{\lambda(0,q)}\left[\frac{\eta(\ell(t, z(t)))}{\ell(t, z(t))}\right],$$

where we write $E_q[\cdot] \equiv E[\cdot|q(0) = q]$ and $E_z[\cdot] \equiv E[\cdot|z(0) = z]$. Use Lemma 4, the definitions (6) of ℓ and λ, and $D_r\lambda(t, r) = 1/(r(1-r))$, to express this expectation as

$$\begin{aligned} E_q[\eta(q(t))] &= \frac{q}{\sqrt{t}}\int_{-\infty}^{\infty}\phi\left(\frac{z(t) - \lambda(0, q)}{\sqrt{t}}\right)\left(e^{-\frac{1}{2}t - z(t)} + 1\right)\eta\left(\frac{1}{e^{-\frac{1}{2}t - z(t)} + 1}\right)dz(t) \\ &= \frac{q}{\sqrt{t}}\int_0^1 \phi\left(\frac{\lambda(t, r) - \lambda(0, q)}{\sqrt{t}}\right)\frac{\eta(r)}{r}\frac{\partial\lambda(t, r)}{\partial r}\,dr \\ &= \frac{q}{\sqrt{t}}\int_0^1 \frac{1}{r^2(1-r)}\phi\left(\frac{1}{\sqrt{t}}\log\left(\frac{r(1-q)}{q(1-r)}\right) - \frac{\sqrt{t}}{2}\right)\eta(r)\,dr. \end{aligned}$$

Using the definition of $L(r, q)$, this yields (8), if we equate the coefficient of $\eta(t)$. □

In particular, this yields a critical time derivative:

Lemma 6 (*t-Derivative*). *The belief transition pdf $q(t)$ satisfies for* $0 < q, r < 1$ *and* $t > 0$:

$$\xi_t(t, q, r) = \xi(t, q, r)\left[-\frac{1}{8} + \frac{L^2(r, q)}{2t^2} - \frac{1}{2t}\right].$$

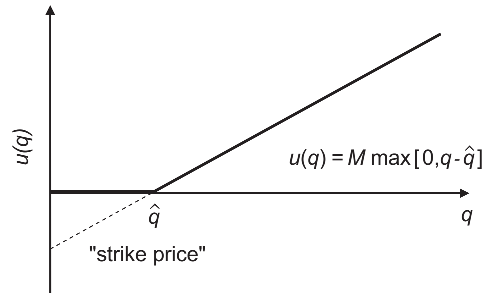

Fig. 3. Information valuation and option pricing.

4. The value of information

4.1. Option pricing analogy

Before deriving our value function, we motivate this with a related exercise done in finance. To this end, simplify matters by positing that action A is a safe action yielding zero payoff, so that $\pi_A^H = \pi_A^L = 0$. Then Eq. (1) can be written as

$$u(q) = M \max\langle 0, q - \hat{q}\rangle,$$

as in Fig. 3. Here we see that $u(q)/M$ can be interpreted as the payoff of a European call option with strike price $\hat{q}$ and underlying asset price q.

Black and Scholes [1] derive the option pricing formula when the underlying asset follows a geometric Brownian motion. They use an arbitrage argument to deduce a parabolic PDE that reduces to the heat equation after a change of variable, using the time to maturity. But geometric Brownian motion is still far more tractable than our nonlinear belief diffusion (2), and thus only a time rescaling of the range variable is needed. By contrast, we use a more complicated transformation.

Refs. [4] and [5] later derived the option pricing formula via martingale methods. The z-process is a martingale under the measure Q just as the discounted asset price is a martingale under the 'pricing measure' in [4]. Our first approach follows this line of thought, but its execution requires a simultaneous range and domain transformation.

4.2. Value function derivation

The expected value $u(t, q)$ of a "quantity'' t of information obeys $u(t, q) = E_q[u(q(t))]$. By the backward equation, $u_t\, dt = E[dq]u_q + \frac{1}{2}E[(dq)^2]u_{qq}$. Since $E[dq] = 0$ by (2):

Lemma 7 (*Inhomogeneous heat equation*). *Expected payoffs $u(t, q)$ satisfy*

$$u_t(t, q) = \tfrac{1}{2}q^2(1 - q)^2 u_{qq}(t, q) \tag{9}$$

for all $(t, q) \in (0, \infty) \times (0, 1)$ with the initial condition[4] $: u(0, q) = u(q)$.

[4] For any $t > 0$, the expected value $u(t, q)$ is a convolution of the V-shaped boundary function $u(0, q)$ and a smooth kernel density. Hence, it is differentiable.

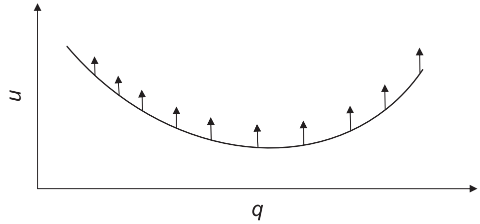

Fig. 4. Analogy with Fourier's law. Fourier's Law of Heat Conduction is shown—i.e., how the heat flow is locally positive exactly when the heat distribution is locally convex on the bar. Specifically, if $u(t, q)$ is the temperature at position q on a bar, with ends held constant at temperatures $u(0)$ and $u(1)$, respectively, then the temperature is increasing $u_t > 0$ iff $u_{qq} > 0$. This is exactly analogous to the behavior of expected payoffs as we acquire more information.

Neatly, (9) asserts the equivalence of two truths: the marginal value of information is non-negative ($u_t \geqslant 0$) and the value of information is convex in beliefs ($u_{qq} \geqslant 0$). Likewise for the heat equation, the temperature gradient within a finite bar obeys a qualitatively similar law where convexity is critical, as Fig. 4 depicts.[5] The option value is convex in the price, just as expected payoffs are convex in beliefs.

As the inhomogeneous heat equation (9) is not directly soluble, we define $z = \log\left(\frac{q}{1-q}\right)$ and transform expected payoffs as $h(t, z) = u\left(t, \frac{1}{1+e^{-z}}\right)(1 + e^{-z})$.

Lemma 8 (*A stochastic representation*). *Expected payoffs are represented as*

$$h(t, z) = E_z^Q\left[\left(e^{-\frac{1}{2}t-z(t)} + 1\right) u\left(\frac{1}{e^{-\frac{1}{2}t-z(t)} + 1}\right)\right]. \tag{10}$$

Proof 1 (*The Martingale method*). We follow [4], and exploit our martingale $z(\cdot)$. The posterior expected payoff $u(t, q)$ equals

$$qE_q^P\left[\frac{q(t)}{q}\frac{u(q(t))}{q(t)}\right] = qE_q^Q\left[\frac{u(q(t))}{q(t)}\right] = qE_{\lambda(0,q)}^Q\left[\left(e^{-\frac{1}{2}t-z(t)}+1\right) u\left(\frac{1}{e^{-\frac{1}{2}t-z(t)}+1}\right)\right]$$

using Lemma 4. Finally, $u(t, q) = qh\,(t, \lambda(0, q))$ gives the desired equation (10).

Proof 2 (*The heat equation*). We now adapt [1]. Change variables[6] in (9) from beliefs q to $Z = \log[q/(1-q)] + t/2 = \lambda(-t, q)$—where t is the elapse time. Let $H(t, Z) = h(t, Z - t/2) = u\left(t, \frac{1}{e^{t/2-Z}+1}\right)\left(e^{t/2-Z} + 1\right)$. Then the heat equation obtains, $H_t(t, Z) = \frac{1}{2}H_{ZZ}(t, Z)$, as we argue in the appendix.

The heat equation solution (e.g. [6, p. 254]) yields:

$$H(t, Z) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{t}}\phi\left(\frac{y}{\sqrt{t}}\right)\left(e^{-Z-y} + 1\right) u\left(\frac{1}{e^{-Z-y} + 1}\right) dy.$$

Finally, using $H(t, Z + t/2) = h(t, Z)$ gives the representation (10). □

---

[5] Shannon [13] also offered a powerful analogy between information and energy.

[6] We thank Robert Israel of UBC for first pointing out a related transformation.

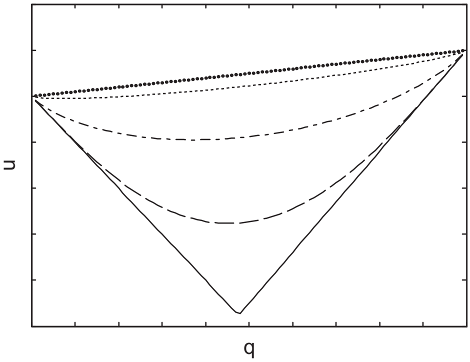

Fig. 5. Posterior expected payoff with different information levels. The parameters are: $\hat{q} = 0.48$, $\pi_A^H = 1$, $\pi_A^L = 2$, $\pi_B^H = 2.1$, and $\pi_B^L = 1$. The bottom solid line graph is $u(q)$, while the top dotted line is $\bar{u}(q)$. Between are $u(1, q)$, $u(5, q)$, $u(15, q)$, the graphs –, – ·,—.

We now exploit this stochastic representation to derive the value of information. By Lemma 1, the long-run limit of the expected payoff is given by

$$\lim_{t \uparrow \infty} u(t, q) = qu(1) + (1 - q)u(0) \equiv \bar{u}(q), \tag{11}$$

i.e., we can write $u(\infty, q) = \bar{u}(q)$. Let us define the *full information gap* as follows

$$\mathrm{FIG}(t, q) = u(\infty, q) - u(t, q) = \bar{u}(q) - u(t, q). \tag{12}$$

Thus, FIG is the difference between the expected payoffs with full information and time t information. We now explore the behavior of the value function $v(t, q)$.

Theorem 1 (*The value of information formula*). *The expected payoff satisfies* $u(t, q) = q\pi_B^H + (1 - q)\pi_A^L - FIG(t, q)$, *where the full information gap* $FIG(t, q)$ *equals*

$$q\left(\pi_B^H - \pi_A^H\right)\Phi\left(-\tfrac{1}{2}\sqrt{t} + \tfrac{1}{\sqrt{t}}L(\hat{q}, q)\right) - (1 - q)\left(\pi_A^L - \pi_B^L\right)\Phi\left(-\tfrac{1}{2}\sqrt{t} - \tfrac{1}{\sqrt{t}}L(\hat{q}, q)\right)$$

and Φ *is the standard normal cdf. The value of information* $v(t, q) = u(t, q) - u(q)$ is

$$v(t, q) = \begin{cases} q\left(\pi_B^H - \pi_A^H\right)\Phi\left(\tfrac{1}{2}\sqrt{t} - \tfrac{1}{\sqrt{t}}L(\hat{q}, q)\right) \\ \quad -(1 - q)\left(\pi_A^L - \pi_B^L\right)\Phi\left(-\tfrac{1}{2}\sqrt{t} - \tfrac{1}{\sqrt{t}}L(\hat{q}, q)\right) & \forall q \leqslant \hat{q}, \\ -q\left(\pi_B^H - \pi_A^H\right)\Phi\left(-\tfrac{1}{2}\sqrt{t} + \tfrac{1}{\sqrt{t}}L(\hat{q}, q)\right) \\ \quad +(1 - q)\left(\pi_A^L - \pi_B^L\right)\Phi\left(\tfrac{1}{2}\sqrt{t} + \tfrac{1}{\sqrt{t}}L(\hat{q}, q)\right) & \forall q \geqslant \hat{q}. \end{cases} \tag{13}$$

The appendix proof uses the log-normal distribution of $e^{-t/2 - z(t)}$ in Lemma 8.

Fig. 5 illustrates the expected payoff with different information levels. The value of information $v(t, q)$ is high for q near the cross-over belief $\hat{q}$, and zero when $q = 0, 1$. The value behaves like the time value of the option, and is increasing in t just as the time value of the option is increasing in the time to maturity.

5. The marginal value of information

Information only has positive marginal value if we hit the cross-over belief $\hat{q}$ at time t; otherwise, prior beliefs dictate the ex post optimal action. The value of information is then increasing in the payoff slope difference $M - m$ and belief variance $\hat{q}^2(1-\hat{q})^2$. Since $v(t, q) = u(t, q) - u(q)$, all time derivatives of the value of information $v(t, q)$ and expected payoffs $u(t, q)$ coincide, and admit a similar expression. Indeed:

Theorem 2 (*Time derivatives*). *The t-derivatives of $v(t, q)$, for $n = 1, 2, \dots,$ are*

$$\left(\frac{\partial}{\partial t}\right)^n v(t,q) = \frac{1}{2}\hat{q}^2(1-\hat{q})^2(M-m)\left(\frac{\partial}{\partial t}\right)^{n-1} \xi(t,q,\hat{q}) \tag{14}$$

for all $(t, q) \in (0,\infty) \times (0, 1)$, where $m \equiv \pi_A^H - \pi_A^L < \pi_B^H - \pi_B^L \equiv M$. In particular, the marginal value of information is given by a scaled standard normal density:

$$v_t(t,q) = \frac{(M-m)q(1-\hat{q})}{2\sqrt{t}}\phi\left(-\frac{1}{2}\sqrt{t} + \frac{1}{\sqrt{t}}L(\hat{q},q)\right). \tag{15}$$

Eq. (15) is proven by differentiating (13). For a less mechanical development of the marginal value of information, write $u_t(t,q) \approx E_q[u_t(\varepsilon, q(t))]$, for small $\varepsilon > 0$. Since the backward equation $u_t = \frac{1}{2}q^2(1-q)^2 u_{qq}$ applies at time $\varepsilon > 0$, we have

$$u_t(t,q) \approx \int_0^1 \frac{1}{2}r^2(1-r)^2 u_{qq}(\varepsilon, r)\xi(t,q,r)\,dr \approx \frac{M-m}{4\varepsilon}\int_{\hat{q}-\varepsilon}^{\hat{q}+\varepsilon} r^2(1-r)^2\xi(t,q,r)\,dr$$

where we approximate u_{qq} near the kink with $u_{qq}(\varepsilon, q) \approx \frac{M-m}{2\varepsilon}$ for all $q \in (\hat{q}-\varepsilon, \hat{q}+\varepsilon)$, and otherwise $u_{qq}(\varepsilon, q) = 0$. Taking the limit $\varepsilon \to 0$, we get $u_{qq}(0,\hat{q}) = \infty$ and otherwise $u_{qq}(0, q) = 0$. Thus, $u_t(t,q) \approx \frac{1}{2}(M-m)\hat{q}^2(1-\hat{q})^2\xi(t,q,\hat{q})$.

Theorem 2 is key, as the demand and price elasticity, respectively, turn on ratios of the value and marginal value to their derivatives, yielding polynomials in $1/t$.

Also, while the value of information behaves continuously as beliefs q converge upon the cross-over belief $\hat{q}$, the marginal value explodes for small t. Indeed:

Corollary 1 (*Derivatives*). *The marginal value of information obeys*[7] :

(a) *For all $t \in (0,\infty)$, $v_t(t,q) \in (0,\infty)$ for all $q \in (0,1)$, while $v_t(\infty, q) = 0$ for all q,*
(b) (*Radner–Stiglitz* [12]) *$v_t(0+, q) = 0$ for all $q \neq \hat{q}$, while $v_t(0+,\hat{q}) = \infty$,*
(c) *$v_{t^n}(0+, q) = 0$ for all $q \neq \hat{q}$ and $n = 2, 3, \dots,$ while $v_{t^n}(0+,\hat{q}) = \infty$.*

The proof is in the appendix. Part (b) is the 'nonconcavity in the value of information' conclusion of [12] and [2], since a marginal value that starts at zero cannot globally fall. (See Fig. 6.) We go beyond this conclusion in part (c), finding that all higher order derivatives also initially vanish for our informational framework.

The Inada condition of [12] or of [2] (their assumption A1) does not apply. Our signal X has mean $\mu_0 t$ and variance t. The Inada in [12] condition fails, and assumption A1 in [2] fails, since

[7] As usual, $v_t(0+, q) = \lim_{s\downarrow 0} v_t(s, q)$, and $v_{t^n}(0+, q) = \lim_{s\downarrow 0} v_{t^n}(s, q)$.

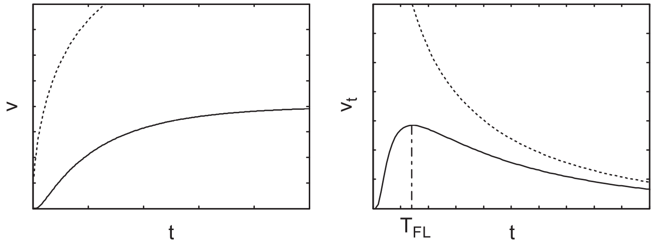

Fig. 6. Value and marginal value of information. At left (right) is the graph of the value (marginal value) of information for the parameter values $\hat{q} = 0.5$, $\pi_H^A = 1$, $\pi_L^A = 2$, $\pi_H^B = 2$, and $\pi_L^B = 1$. The solid lines are the values for $q = 0.2$ or 0.8, where the nonconcavity arises, and the dotted lines correspond to the values for the cross-over belief $q = 0.5$.

the signal density derivative in t explodes at $t = 0$—as [2] acknowledge for [7]. But had we assumed the convex continuous action model of [7], nonconcavity would be no longer clear: indeed, $u_t(0+, q) = \lim_{t\to 0} u_t(t, q) = \lim_{t\to 0} \frac{1}{2}q^2(1-q)^2 u_{qq}(t, q) = \frac{1}{2}q^2(1-q)^2 u_{qq}(q) > 0$ by assumption. [8] As standard as our Gaussian model is—the limit of the discrete sampling models, as we see in Section 8—it escapes known sufficient conditions. And yet $v_t(0+, q) = 0$.

We now globally describe the marginal value of information. Let us define an in*fl*ection point in the value of information, where the marginal value peaks:

$$T_{\mathrm{FL}}(q) = 2\left[\sqrt{1 + L^2(\hat{q}, q)} - 1\right]. \tag{16}$$

This inflection point demand is surprisingly independent of the payoff levels except insofar as it affects the cross-over belief $\hat{q}$. This holds in spite of the fact that the marginal value of information (15) is indeed increasing in the payoff stakes $M - m$.

Theorem 3 (*Value of information*). *Fix $q \in (0, 1)$ with $q \neq \hat{q}$.*

(a) *The value of information is convex until $t = T_{\mathrm{FL}}(q)$, after which it is concave.*
(b) *The marginal value is rising until $T_{\mathrm{FL}}(q)$ then falling. It is convex in $[0, T_1(q)]$, concave in $(T_1(q), T_2(q))$, and convex for $[T_2(q), \infty)$, where $T_{\mathrm{FL}}(q) \in (T_1(q), T_2(q))$.*

Proof. Eq. (16) yields $T_{\mathrm{FL}}(q) \in [0, \infty)$ and $T_{\mathrm{FL}}(q) = 0$ iff $q = \hat{q}$. From (9) and Theorem 2,

$$v_{tt}(t, q) = v_t(t, q)\left[-\frac{1}{8} + \frac{L^2(\hat{q}, q)}{2t^2} - \frac{1}{2t}\right], \tag{17}$$

where $v_t(t, q) > 0$, by Corollary 1. Now, $v_{tt}(t, q) = 0$ gives $\Upsilon(t, q) \equiv t^2 + 4t - 4L^2(\hat{q}, q) = 0$, yielding (16). Part (a) owes to $\Upsilon(t, q) \lesseqgtr 0$ for all $t \lesseqgtr T_{\mathrm{FL}}(q)$.

For (b), note that $v_{ttt}(0+, q) = 0$ according to Corollary 1. Second, by Theorem 2

$$v_{ttt}(t, q) = v_t(t, q)\left[\left(-\frac{1}{8} + \frac{L^2(\hat{q}, q)}{2t^2} - \frac{1}{2t}\right)^2 - \frac{L^2(\hat{q}, q)}{t^3} + \frac{1}{2t^2}\right] \tag{18}$$

and hence $v_{ttt}(t, q) = 0$ if $t^4 + 8t^3 + (48 - 8L^2)t^2 - 96L^2 t + 16L^2 = 0$. Clearly, $v_{ttt} > 0$ near $t = 0$. So if there is only one positive root $T_1(q)$ then $v_{ttt}(t, q) < 0$ for all $t > T_1(q)$.

[8] In work in progress on the non-concavity property in [12], we explore this continuum action world more fully.

Since $v_{ttt}(t, q) > 0$ for large t, if there is a strictly positive root then there must be two strictly positive roots $T_1(q)$ and $T_2(q)$. If there are no positive roots then $v_{ttt}(t, q) > 0$ for all $t > 0$. By Corollary 1(c), this gives $v_{tt}(t, q) > 0$ for all $t > 0$—contradicting part (a). So there are two positive roots. We have $T_1(q) < T_{\text{FL}} < T_2(q)$ since $v_{tt}(T_{\text{FL}}, q) = 0$ and $v_{tt}(t, q) > 0$ for $t < T_{\text{FL}}$, hence (b). □

By Lemma 6, the convexity before $T_{\text{FL}}(q)$ owes to the rising transition pdf ($\xi_t(t, q, \hat{q}) \geqslant 0$) and the concavity after $T_{\text{FL}}(q)$ to the decreasing pdf ($\xi_t(t, q, \hat{q}) \leqslant 0$).

6. The demand for information

6.1. The demand curve

We now consider linear pricing of information $c(t) = ct$, where c is a strictly positive constant. *Demand* $\tau(c, q)$ maximizes consumer surplus

$$\Pi(t, q) = u(t, q) - ct = u(q) + v(t, q) - ct. \tag{19}$$

We fix $q \neq \hat{q}$, ignoring the cross-over belief $\hat{q}$, nongeneric since it is a single point; we can thus avoid hedging our theorems. Because of the non-concavity near quantity 0, and since the marginal value finitely peaks, there exists a *choke-off cost* $c_{\text{CH}}(q) > 0$, above which demand is zero, and an implied least *choke-off demand*, $T_{\text{CH}}(q) > 0$. At the cost $c_{\text{CH}}(q) > 0$, demand is $T_{\text{CH}}(q)$ and consumer surplus is zero. Thus, marginal value is falling, and so $T_{\text{CH}}(q) \geqslant T_{\text{FL}}(q)$, as in Fig. 7. Summarizing:

$$\text{Cost 'choke-off'}: \quad c_{\text{CH}}(q) = v_t(T_{\text{CH}}(q), q) = \frac{v(T_{\text{CH}}(q), q)}{T_{\text{CH}}(q)}. \tag{20}$$

Define $T_{\text{FOC}}(c, q) > T_{\text{FL}}(q)$ by the FOC $v_t(T_{\text{FOC}}(c, q), q) = c$. This is well-defined iff $c \leqslant v_t(T_{\text{FL}}(q), q)$, since $v_{tt}(T_{\text{FL}}(q), q) < 0$ on $(T_{\text{FL}}(q), \infty)$ (Theorem 3(b)). The FOC captures demand precisely when the cost falls below the choke-off cost.

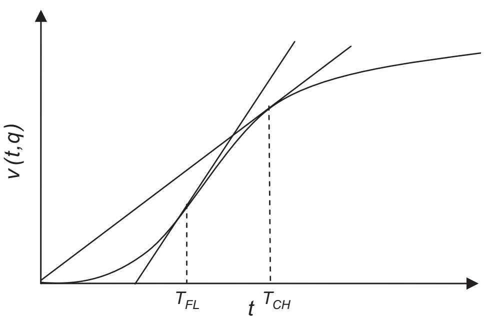

Fig. 7. The information non-concavity. The choke-off demand $T_{\text{CH}}(q)$ exceeds the peak marginal value demand $T_{\text{FL}}(q)$ due to the non-concavity of information. The demand curve is not simply the falling portion of the marginal value of information.

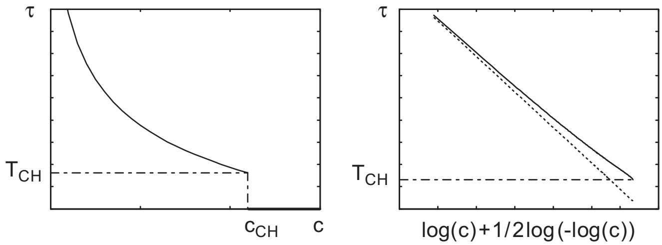

Fig. 8. Optimal and approximate large demand. The true demand curve is depicted in the left figure. In the right figure, true demand is the solid line and the approximation is the dotted line. Parameter values: $q = 0.2$ or 0.8, $\hat{q} = 0.5$, $\pi_H^A = 1$, $\pi_L^A = 2$, $\pi_H^B = 2$, and $\pi_L^B = 1$.

Theorem 4. *$\tau(c, q) = 0$ if $c > c_{\mathrm{CH}}(q)$ and $\tau(c, q) = T_{\mathrm{FOC}}(c, q)$ if $c \leqslant c_{\mathrm{CH}}(q)$.*

Proof. Observe that by (19), $\Pi(t, q) = \int_0^t [v_t(s, q) - c]\, ds$. The integrand is first negative, since $v_t(0+, q) = 0$, and eventually negative, since $v_t(\infty, q) = 0$. If $c > c_{\mathrm{CH}}(q)$, then the integral (consumer surplus) is always negative, and so the optimal demand is $t = 0$. Otherwise, if $c \leqslant c_{\mathrm{CH}}(q) < v_t(T_{\mathrm{FL}}(q), q)$, then $T_{\mathrm{FOC}}(c, q)$ exists, and by Theorem 3, the integrand is positive for an interior interval ending at $T_{\mathrm{FOC}}(c, q)$, where $v_t(T_{\mathrm{FOC}}(c, q), q) - c = 0$. Thus, the integral is maximized at $t = T_{\mathrm{FOC}}(q)$. □

Corollary 2 (*Law of demand*). *Demand is falling in the price c, for $c < c_{\mathrm{CH}}(q)$.*

Indeed, simply apply $v_{tt}(t, q) < 0$ for all $t > T_{\mathrm{FL}}(q)$ (true by Theorem 3).[9] The law of demand applies to information too, but only after the price drops below the choke-off level $c_{\mathrm{CH}}(q)$, warranting positive demand. Fig. 8 illustrates these results: the jump in information demand as costs drop, as well as the Law of Demand.

6.2. *Quantifying the lumpiness*

We now explore the size of the nonconcavity in the demand for information. The most direct approach here is to quantify the minimum expenditure $T v_t(T, q)$ on information that the *DM* must incur. Of course, this amount should increase in the maximum payoff stakes, simply because the marginal value of information does, by Theorem 2. Additionally, if beliefs are near 0 or 1, then information demand vanishes. Seeking an appropriate normalization, let the *expected payoff stakes* denote the maximum expected payoff loss from choosing a wrong action. The worst case scenario occurs at the cross-over belief $q = \hat{q}$, with expected payoffs $\hat{q}$[max loss if $\theta = H$] + $(1 - \hat{q})$[max loss if $\theta = L$], which equals:

$$\hat{q}[\pi_B^H - \pi_A^H] + (1 - \hat{q})[\pi_A^L - \pi_B^L] = \hat{q}(1 - \hat{q})(M - m).$$

This vanishes when $\hat{q} = 0, 1$, and increases in the maximum payoff stakes $M - m$.

[9] This result owes to the supermodularity of $v(t) - tc$ in $(t, -c)$. We thank Ed Schlee for this.

Theorem 5 (*Least positive information costs*). *The average lower bound on information expenditures normalized by the payoff stakes exceeds* 0.025, *or*

$$\int_0^1 \frac{T_{\rm CH}(r)v_t(T_{\rm CH}(r),r)}{\hat{q}(1-\hat{q})(M-m)}\,dr > 0.025. \tag{21}$$

Proof. Suppressing the arguments of $L = L(\hat{q}, q)$ and $T_{\rm CH}(q)$, we have

$$\frac{v(T_{\rm CH},q)}{\hat{q}(1-\hat{q})(M-m)} = \frac{\int_0^{T_{\rm CH}} v_t(s,q)\,ds}{\hat{q}(1-\hat{q})(M-m)} = \frac{1}{2}\sqrt{\frac{q(1-q)}{2\pi\hat{q}(1-\hat{q})}}\int_0^{T_{\rm CH}} \frac{1}{\sqrt{s}}\exp\left(-\frac{s}{8}-\frac{L^2}{2s}\right)ds.$$

Using this equation, a lower bound on (21) is 0.025, as we show in the appendix. □

We take an average here because the threshold choke-off cost $c_{\rm CH}(q)$ vanishes as beliefs q approach $\hat{q}$ or the extremes 0, 1. Thus, the minimum information purchase likewise vanishes nearing those three beliefs, and only an average makes sense.

6.3. Demand as a function of beliefs

A classic question asked of Bayesian sequential learning models is the range of beliefs with positive experimentation.

Theorem 6 (*Interval demand*). *Demand* $\tau(c,q) > 0$ *iff beliefs* q *belong to an interior interval* $\left(\underline{q}(c), \bar{q}(c)\right)$, *where the thresholds* $0 < \underline{q}(c) < \hat{q} < \bar{q}(c) < 1$ *obey*

$$v\left(\tau\left(c,\underline{q}(c)\right),\underline{q}(c)\right) = \tau\left(c,\underline{q}(c)\right)c \quad \text{and} \quad v\left(\tau(c,\bar{q}(c))\right) = \tau\left(c,\bar{q}(c)\right)c. \tag{22}$$

Also, the choke-off demands are $T_{\rm CH}(\bar{q}(c)) = \tau(c,\bar{q}(c))$ *and* $T_{\rm CH}(\underline{q}(c)) = \tau(c,\underline{q}(c))$.

Proof. Demand is positive at belief $\hat{q}$, since $v_t(0+,\hat{q}) = \infty$, by (b) in Corollary 1. Also, demand vanishes at $q = 0, 1$, since $v_t(t,0) = u_t(t,0) = 0$ and $v_t(t,1) = u_t(t,1) = 0$ for all t. Thus, any interval structure obeys $0 < \underline{q}(c) < \hat{q} < \bar{q}(c) < 1$.

Next, when positive, demand obeys the FOC $v_t(\tau(q),q) = c$. While it suffices to prove that $v_t(t,q)$ is strictly quasi-concave in q, we instead show local quasi-concavity at the optimal demand $\tau(c,q)$—i.e., $v_{tq}(\tau(c,q),q) = 0$ implies $v_{tqq}(\tau(c,q),q) < 0$.

Differentiating the FOC yields $\tau_q(c,q) = -v_{tq}/v_{tt}$, if we suppress arguments. So,

$$\tau_{qq}(c,q) = -\frac{1}{v_{tt}}\left(v_{tqq} + v_{ttq}\tau_q\right) + \frac{v_{tq}}{v_{tt}^2}\left(v_{ttq} + v_{ttt}\tau_q\right) = -\frac{1}{v_{tt}}v_{tqq} \tag{23}$$

because our premise $v_{tq}(\tau(c,q),q) = 0$ is equivalent to $\tau_q(c,q) = 0$, by the FOC. By Theorem 3(b), we have $v_{tt}(\tau(c,q),q) < 0$, and so $v_{tqq}(\tau(c,q),q)$ and $\tau_{qq}(c,q)$ share the same sign. Since $u_{tt}(\tau(c,q),q) = \frac{1}{2}q^2(1-q)^2u_{tqq}(\tau(c,q),q)$ by Lemma 7,

$$-\frac{v_{tqq}(\tau(c,q),q)}{v_{tt}(\tau(c,q),q)} = -\frac{u_{tqq}(\tau(c,q),q)}{u_{tt}(\tau(c,q),q)} = -\frac{2}{q^2(1-q)^2} < 0. \tag{24}$$

Finally, demand vanishes when the *DM* is indifferent between buying and not buying at all—i.e., at the choke-off level. So (22) follows, and $T_{\rm CH}(q)$ is as described. □

These results strikingly differ from their analogs in a dynamic setting. That the information demand is positive precisely on an interior interval is in harmony with the standard sequential experimentation result (see [10]). But it holds for an entirely unrelated reason! In sequential experimentation, the *DM* stops when his *marginal* costs and benefits of more experimentation balance. In our static demand setting, the *DM* stops buying information when *total* costs and benefits of any information purchase balance. So given the nonconcavity in the value of information, this demand choke-off decision turns on considerations of *global optimality.*

The next theorem gives the relationship between positive demand and beliefs.

Theorem 7 (*Hill-shaped demand*). *$\tau(c, \cdot)$ is quasi-concave for $q \in \left(\underline{q}(c), \bar{q}(c)\right)$.*

Proof. By Theorem 6, $\tau(c, q) > 0$. So $\tau_q = 0$ implies $\tau_{qq} < 0$ by (23)–(24). □

A comparison with the dynamic case is instructive, and here we contrast the result, and not just its logic. [10] assume a convex cost of information in a sequential experimentation model and deduce instead that information demand is U-shaped and convex, and not hill-shaped and concave. The static demand solution is the intuitive one, with demand greatest when the *DM* is most uncertain.

6.4. The elasticity of demand

The elasticity of the demand is $|\tau_c(c, q)c/\tau(c, q)|$. When the elasticity equals 1, the demand level is $T_E(q)$, and revenue $v_t(t, q)t$ is maximized. By this fact, we characterize $T_E(q)$ below, via the belief derivative (17):

$$T_E(q) = -\frac{v_t\left(T_E(q), q\right)}{v_{tt}\left(T_E(q), q\right)} = 2\left[1 + \sqrt{1 + L^2(\hat{q}, q)}\right] = T_{\mathrm{FL}}(q) + 4. \tag{25}$$

Like the peak marginal value $T_{\mathrm{FL}}(q)$, the unit elastic demand does not depend on the underlying payoff stakes, apart from the dependence on the cross-over belief $\hat{q}$. Further, the marginal value is clearly falling at $T_E(q)$, since it exceeds $T_{\mathrm{FL}}(q)$. We show that it lies above the choke-off demand if the belief is sufficiently interior.

Theorem 8. (a) *Demand is initially elastic for $q \in (\acute{q}, \grave{q})$, where $0 < \acute{q} < \hat{q} < \grave{q} < 1$.*
(b) *Demand elasticity is decreasing in the cost c, for all $c \leqslant c_{\mathrm{CH}}(q)$.*

Observe that $(\acute{q}, \grave{q}) \subset (\underline{q}(c), \bar{q}(c))$ as demand is positive for $q \in (\acute{q}, \grave{q})$. Let $c_{\mathrm{E}}(q) = v_t\left(T_E(q), q\right)$, the cost with unit demand elasticity. Then $c_{\mathrm{E}}(q) < c_{\mathrm{CH}}(q)$ iff $q \in (\acute{q}, \grave{q})$. Information demand is elastic for $c \in (c_{\mathrm{E}}(q), c_{\mathrm{CH}}(q)]$ and inelastic if $c < c_{\mathrm{E}}(q)$. The marginal value of information drops off so fast given the exponentially thin Gaussian tail that the demand elasticity falls monotonically.

6.5. Large demand

We now derive a simple asymptotically applicable demand formula. It is consistent with the one [9] derived using large deviation theory for conditionally iid samples from any signal distribution.

Our work here follows in a Gaussian framework and so is more refined: we specify an additional error term, [10] and show that it is positive—i.e., the limit demand curve is approached from above.

Theorem 9 (*Low prices*). *If c is small then the optimal demand is given by*

$$\tau(c,q)/8 = F(q) - \log(c) - \frac{1}{2}\log(-\log(c)) + \frac{\log(-\log(c))}{4\log(c)}(1+o(1)), \quad (26)$$

where $o(1)$ vanishes in c, and where $F(q) = \frac{1}{2}\log[q(1-q)\hat{q}(1-\hat{q})/64\pi] + \log(M-m)$.

Observe that the approximate difference $2[\log(-\log(c))]/\log(c)$ between demand and $F(q) - \log c - \frac{1}{2}\log(-\log(c))$ is negative, and vanishing in c (see Fig. 8). Therefore, the three c-dependent terms of the demand function (26) provide (in order, adding them from left to right) increasingly accurate approximations. As the cost of information vanishes, demand is essentially logarithmic.

Proof of Theorem 9. If the cost c is small then $T_{\rm FOC}(c,q)$ exists and $u(T_{\rm FOC}(c,q),q) \geqslant u(0,q)$. So in this case $\tau(c,q) = T_{\rm FOC}(c,q)$. Second, from the FOC $v_t(\tau(c,q),q) = c$:

$$c = \frac{(M-m)q(1-\hat{q})}{2\sqrt{2\pi\tau(c,q)}}\exp\left\{-\frac{1}{2}\left[\frac{1}{4}\tau(c,q) - L(\hat{q},q) + \frac{L^2(\hat{q},q)}{\tau(c,q)}\right]\right\}.$$

Taking logs, the definitions of $L(\hat{q},q)$ and $F(q)$ yield the *log inverse demand curve*:

$$\log(c) = F(q) - \frac{1}{2}\log(\tau(c,q)/8) - \frac{L^2(\hat{q},q)}{16\tau(c,q)/8} - \frac{1}{8}\tau(c,q) = F(q) - \psi(\tau(c,q)/8), \quad (27)$$

where $\psi(x) = x + \frac{1}{2}\log x + B/x$, for $B = L^2(\hat{q},q)/16$. For large x, we have [11] $\psi'(x) = 1 + 1/(2x) - B/x^2 > 0$, so that ψ^{-1} exists. So (27) yields the demand curve $\tau(c,q) = 8\psi^{-1}(F(q) - \log c)$. The appendix shows that $\psi^{-1}(x) = x - \frac{1}{2}\log x + \frac{1}{4}[(\log x)/x](1+o(1))$ and $\psi^{-1}(x) > x - \frac{1}{2}\log x$. Then (26) follows as $\tau(c,q)/8$ equals

$$F(q) - \log(c) - \frac{1}{2}\log(-\log(c)) - \frac{1}{2}\log\left(1 - \frac{F(q)}{\log(c)}\right) + \frac{1}{4}\frac{\log(F(q) - \log(c))}{F(q) - \log(c)}(1+o(1)). \qquad \square$$

Recall our normalization $\zeta = 1$. More generally, the demand function in (26) must be divided by ζ^2. Ref. [9] analyze the large demand for information as the price c vanishes, for any arbitrary signal, not just for weak or gaussian signals as we do here. They define a general information index for a signal, $\rho \in (0,1)$, where 0 means perfectly informative, and 1 uninformative. They show that the demand function for small c has the same log form as in (26), with same slope when $\zeta^2 = -8\log(\rho)$:

$$\frac{\tau(c,q)}{8\log(\rho)} = F(q) - \left[\log(c) + \frac{1}{2}\log(-\log(c))\right] + \frac{\log(-\log(c))}{4\log(c)}(1+o(1)).$$

[10] In the discrete signal world of [9], their formula was eventually accurate within one signal.

[11] This holds for $x > 1 + \sqrt{1 + L^2(\hat{q},q)/16}$, which is less than $T_E(q)$. In other words, certainly starting when demand is inelastic, our inverse demand curve $\tau(c,q) = 8\psi^{-1}(F(q) - \log(c))$ is valid.

At small prices, the optimal i.i.d. sample size of any signal with index ρ is approximately the demand time of a Brownian motion with squared signal-to-noise ratio $\zeta^2 = -8\log(\rho)$. This rises from 0 to ∞ as informativeness rises (ρ falls from 1 to 0).

7. 'Small bits' of information

7.1. Beliefs

Seeking a theory of variable quantity information, it surely must be measured in sufficiently small units. We now prove that our theory built on the diffusion process $X(\cdot)$ well-approximates models assuming very 'small bits' of information. Assume the *DM* chooses the number n of i.i.d. draws from any signal, and let the informational content and the cost of each draw jointly vanish. We show that our Gaussian information approximates the value of and optimal number of cheap, weak signals.

Let $\{G(\cdot|\theta,\Delta)\}$ be a simple signal—a family of c.d.f.s, each indexed by the state $\theta \in \{L, H\}$, with measurable support $\mathcal{Z}$ independent of θ. Assume that the signal becomes pure noise as Δ vanishes, as the likelihood ratio $\lambda(Z|\Delta) = dG(Z|L,\Delta)/dG(Z|H,\Delta) > 0$ tends to 1. Here, Δ is the real elapse duration of a time interval in discrete time. In the (continuous) time span $[0, t]$, the *DM* observes $n = \lfloor t/\Delta \rfloor$ draws from $G(\cdot|\theta,\Delta)$ at times $\Delta, 2\Delta, \ldots, n\Delta \doteq t$, where $\lfloor a \rfloor$ is the largest integer at most a. So as Δ vanishes, the *DM* sees an exploding number of increasingly uninformative conditionally i.i.d. signal outcomes at high frequency.

Introduce a sequence $\{Z_n\}$ of conditionally iid random variables drawn from $G(\cdot|\theta,\Delta)$ in state θ. Next, define a belief process q_n^{Δ}, according to Bayes rule:

$$q_n^{\Delta} = \frac{q_{n-1}^{\Delta}}{q_{n-1}^{\Delta} + [1 - q_{n-1}^{\Delta}]\lambda(Z_n|\Delta)}. \tag{28}$$

Extend this to a process on the real line: $q^{\Delta}(t) = q^{\Delta}_{\lfloor t/\Delta \rfloor}$. The appendix proves:

Theorem 10 (*Small bits*). *The discrete Markov process* (28) *converges weakly to the diffusion in* (2), *namely* $q^{\Delta}(\lfloor t/\Delta \rfloor \Delta) \Rightarrow q(t)$, *if for each* $\theta \in \{H, L\}$, *we have*

$$\int \left[1 - \lambda(z|\Delta)\right]^2 dG(z|\theta,\Delta) = \Delta + o(\Delta). \tag{29}$$

Unless the likelihood ratio $\lambda(z|\Delta)$ converges to 1, beliefs will jump and our limit belief process (2) will not be continuous. Theorem 10 implies that $\lambda(z|\Delta) \to 1$ in the mean-square sense—as needed for Ito integrals to converge.

We now illustrate, by way of example, which types of signals satisfy this condition and thus can in fact be approximated by our Gaussian model.

Example 1 (*Garbling fixed signals*). We identify a broad class of examples obeying (29) using *any* signal, whose cdfs are $\{F(\cdot|\theta)\}$ (possibly containing atoms). "Garble" F as follows: in state θ, the signal is drawn from $F(\cdot|\theta)$ with chance $\frac{1}{2} + \sqrt{\Delta}$, and from the "incorrect" distribution with chance $\frac{1}{2} - \sqrt{\Delta}$. For instance,

$$G(z|H,\Delta) = \left(1/2 + \sqrt{\Delta}\right) F(z|H) + \left(1/2 - \sqrt{\Delta}\right) F(z|L).$$

This yields a state-independent support $\mathcal{Z}$. Define the unconditional signal cdf $F(z) = \frac{1}{2}F(z|H) + \frac{1}{2}F(z|L)$, and Radon–Nikodym derivatives $\alpha(z|\theta) = dF(z|\theta)/dF(z)$. Assume:

$$1 = 8 \int_{\mathcal{Z}} \frac{[f(z|H) - f(z|L)]^2}{f(z|H) + f(z|L)} \, dF(z). \tag{30}$$

This fixes the signal–noise ratio at one; more generally, garbling still gives the result, but with a different signal–noise ratio. For since $\lambda(Z|\Delta) = dG(Z|L,\Delta)/dG(Z|H,\Delta)$,

$$\left[1 - \lambda(z|\Delta)\right]^2 = \frac{16[f(z|H) - f(z|L)]^2 \Delta}{\left(f(z|H) + f(z|L) + 2\,[f(z|H) - f(z|L)]\sqrt{\Delta}\right)^2}. \tag{31}$$

Thus, $[1 - \lambda(z|\Delta)]^2 = O(\Delta)$, so that $\lambda(z|\Delta) = 1 + O(\sqrt{\Delta})$. This gives

$$\int_{\mathcal{Z}} \left[1-\lambda(z|\Delta)\right]^2 dG(z|H,\Delta) = \int_{\mathcal{Z}} \frac{8\,[f(z|H)-f(z|L)]^2\,\Delta}{f(z|H)+f(z|L)+2\,[f(z|H)-f(z|L)]\sqrt{\Delta}} \, dF(z).$$

Thus, (29) follows from (30) for the state $\theta = H$. For state L, we have

$$\int_{\mathcal{Z}} \left[1 - \lambda(z|\Delta)\right]^2 \, dG(z|L,\Delta) = \int_{\mathcal{Z}} \lambda(z|\Delta) \left[1 - \lambda(z|\Delta)\right]^2 \, dG(z|H,\Delta).$$

This is $[1 + O(\sqrt{\Delta})][\Delta + o(\Delta)] = \Delta + o(\sqrt{\Delta})$ when $\theta = L$, since $\lambda(z|\Delta) = 1 + O(\sqrt{\Delta})$.

Example 2 (*Low precision Gaussian signals*). Weak Gaussian signals produce a Gaussian signal in the large. Assume that $G(z|H,\Delta)$ and $G(z|L,\Delta)$ are the cdf's of normal variables with mean $\pm\frac{1}{2}$ and variance $1/\Delta$. Then the Radon–Nikodym derivative is $\lambda(z|\Delta) = dG(z|L,\Delta)/dG(z|H,\Delta) = \exp\left(\frac{\Delta}{2}\left[(z-1/2)^2 - (z+1/2)^2\right]\right)$. We can show:

$$\int_{-\infty}^{\infty} \left[1 - \lambda(z|\Delta)\right]^2 \, dG(z|H,\Delta) = e^{\Delta} - 1 = \Delta + o(\Delta),$$

$$\int_{-\infty}^{\infty} \left[1 - \lambda(z|\Delta)\right]^2 \, dG(z|L,\Delta) = 1 - 2e^{\Delta} + e^{3\Delta} = \Delta + o(\Delta).$$

7.2. *Value and demand approximation*

The ex ante value of seeing $n = \lfloor t/\Delta \rfloor$ conditionally i.i.d. draws from $\{G(\cdot|\theta,\Delta)\}$ is

$$v^{\Delta}(t,q) \equiv E_q[u(q^{\Delta}(\lfloor t/\Delta \rfloor \Delta))] - u(q),$$

where the expectation is taken over the distribution of the discrete time belief process $q^{\Delta}(\lfloor t/\Delta \rfloor \Delta)$. Since the latter converges weakly to $q(t)$, and $u(\cdot)$ is continuous, the discrete value function converges to the continuous one for every prior $q \in [0,1]$:

$$v^{\Delta}(t,q) = v(t,q) + o(\Delta). \tag{32}$$

If $v^0(t,q) \equiv v(t,q)$, then $v^{\Delta}(t,q)$ is continuous in Δ at $\Delta = 0$, for every $t \geqslant 0$.

Consider the decision problem where the *DM* can (non-sequentially) purchase n conditionally i.i.d. signals for $nc\Delta$, yielding payoff $\Pi^{\Delta}(n\Delta|c,q) = v^{\Delta}(n\Delta,q) - cn\Delta$. The optimization problem is

$$\sup_{n\in\mathbb{N}} \Pi^{\Delta}(n\Delta|c,q). \tag{33}$$

Since $\Pi^{\Delta}(0|c,q) = v(0,q)$ and $\lim_{n\to\infty} \Pi^{\Delta}(n\Delta|c,q) = -\infty$, a finite non-negative maximum of $\Pi^{\Delta}(n\Delta|c,q)$ over $n = 0, 1, 2, \ldots$ exists—let us call that set $N^{\Delta}(c,q)$.

The analogous continuous time problem when observing $X(t)$ is $\sup_{t\geqslant 0} \Pi(t|c,q)$, where $\Pi(t|c,q) = v(t,q) - ct$. This yields the real demand function $\tau(c,q)$. By (32):

$$\Pi^{\Delta}(t|c,q) = \Pi(t|c,q) + o(\Delta).$$

Theorem 11 (*Demand convergence*). *There is a selection* $n^{\Delta}(c,q) \in N^{\Delta}(c,q)$ *such that for an open and dense set of parameters* $\pi_A^{\theta}, \pi_B^{\theta}, q, c > 0$, *we have*:

$$n^{\Delta}(c,q)\Delta = \tau(c,q) + o(\Delta). \tag{34}$$

The proof is in the appendix. Then Eq. (34) implies that the discrete demand elasticity approximates the continuous demand one.

8. Conclusion

We have measured information with the precision of Gaussian additive noise masking the payoff-relevant state of nature. Here, we have completely and tractably characterized the value of, and demand for, information in the two state world with two actions. *We have focused on the two action case because it most clearly reveals the link to the classic option pricing exercise.* For buying information affords one an option to discretely change one's action, just as an option allows one to buy a single share of a stock. This similarity is the main *conceptual contribution* of this paper, using the logic of the inherently dynamic option pricing problem in our static environment. Indeed, we represent the value of information as the expected payoff of an option on a stochastic belief process. In our key *technical contribution*, we show how to transform the belief process so that the value function can be explicitly solved. We have given two routes to this same target, each tracing a different solution of the option pricing exercise. Our change-of-measure mathematics is essential here for the same reason that it was for the option pricing formula. However, given the value formula, our demand analysis proceeds just using elementary methods.

We then gave the full picture of the famous informational nonconcavity for the first time, graphing the marginal value schedule for Gaussian precision. We prove that this well approximates a large class of "small bit" models of conditionally iid signals. We have also characterized the elasticity of information, the large demand formula, and the dependence of demand on beliefs. Our theory extends, with difficulty, to a model with any finite set of actions, since that merely adds to the number of cross-over beliefs. The restriction to two states is real.

Kihlstrom [7] simply exploited the self-conjugate Gaussian property for his particular payoff function, bi-passing any analysis of the belief process. But a learning model is generally useful insofar as one knows how posterior beliefs evolve stochastically. Indeed, we show that the marginal value of information is proportional to the transition belief density. Either a finite action space or state space invalidates his approach, and one must treat the learning problem seriously. And indeed, many choices in life are inherently discrete, like whether to change jobs or build a prototype. For us, beliefs are not linear in signal realizations, so that solving for this belief density requires new solution methods.

Finally, prompted by a referee, we observe that ours is not simply a model of precision. There are other distributions besides the Gaussian with a state-dependent mean μ_θ and state-independent precision t—for instance, the Gamma distribution. Yet, a Gamma is not approximated by our Gaussian signal.

Acknowledgments

We acknowledge useful suggestions of a referee, as well as Boyan Jovanovic, Xu Meng, Paavo Salminen, Ed Schlee, and Xiqao Xu, and the comments from seminars at Toronto, Georgetown, Michigan, UC - Santa Cruz, the Decentralization Conference (Illinois), the INFORMS Applied Probability Conference (Ottawa), and the SED (Budapest). The NSF funded this research (grant #0550043).

Appendix A. Omitted proofs

A.1. Limit beliefs: Proof of Lemma 1(a)

For all $q \in (0, 1)$, $\int_{q-\varepsilon}^{q+\varepsilon} \frac{2\,dy}{y^2(1-y)^2} < \infty$ for some $\varepsilon > 0$. So Feller's test for explosions [6, Theorem 5.29] implies part (a) because

$$\int_0^c \frac{2\,dq}{q^2(1-q)^2} \geqslant 2\int_0^c \frac{1}{q}\,dq = \infty \quad \forall c \in (0, 1).$$

A.2. The heat equation: completing Proof 2 of Lemma 8

Simply take the derivatives below and apply $u_t = \frac{1}{2}q^2(1-q)^2 u_{qq}$:

$$H_t(t, Z) = \frac{e^{t/2-Z}}{2} u\left(t, \frac{1}{e^{t/2-Z}+1}\right) - \frac{e^{t/2-Z}}{2(e^{t/2-Z}+1)} u_q\left(t, \frac{1}{e^{t/2-Z}+1}\right) + \left(e^{t/2-Z}+1\right) u_t\left(t, \frac{1}{e^{t/2-Z}+1}\right),$$

$$H_{ZZ}(t, Z) = e^{t/2-Z} u\left(t, \frac{1}{e^{t/2-Z}+1}\right) - \frac{e^{t/2-Z}}{e^{t/2-Z}+1} u_q\left(t, \frac{1}{e^{t/2-Z}+1}\right) + \frac{\left(e^{t/2-Z}\right)^2}{\left(e^{t/2-Z}+1\right)^3} u_{qq}\left(t, \frac{1}{e^{t/2-Z}+1}\right).$$

A.3. Value function derivation: Proof of Theorem 1

By Lemma 8, the expected payoff can be represented as an integral as follows:

$$u(t, q) = \int_{-\infty}^{\infty} \left((1-q)\, e^{-\frac{1}{2}t-\sqrt{t}y} + q\right) u\left(\frac{1}{\left(\frac{1}{q}-1\right) e^{-\frac{1}{2}t-\sqrt{t}y}+1}\right) \phi(y)\,dy.$$

Let us exploit symmetry $\phi(y) = \phi(-y)$. Since $u(q) = \max\langle \pi_A^L + mq, \pi_B^L + Mq\rangle$,

$$u(t, q) = q\int_{\hat{y}(t,q)}^{\infty} \left(\left(\frac{1}{q}-1\right) e^{-\frac{1}{2}t+\sqrt{t}y}+1\right) \left(\pi_A^L + \frac{m}{\left(\frac{1}{q}-1\right)e^{-\frac{1}{2}t+\sqrt{t}y}+1}\right) \phi(y)\,dy$$

$$+q\int_{-\infty}^{\hat{y}(t,q)}\left(\left(\frac{1}{q}-1\right)e^{-\frac{1}{2}t+\sqrt{t}y}+1\right)\left(\pi_B^L+\frac{M}{\left(\frac{1}{q}-1\right)e^{-\frac{1}{2}t+\sqrt{t}y}+1}\right)\phi(y)\,dy$$

$$=q\left(\pi_A^L+m\right)\int_{\hat{y}(t,q)}^{\infty}\phi(y)\,dy+(1-q)\pi_A^L\int_{\hat{y}(t,q)}^{\infty}e^{-\frac{1}{2}t+\sqrt{t}y}\phi(y)\,dy$$

$$+q\left(\pi_B^L+M\right)\int_{-\infty}^{\hat{y}(t,q)}\phi(y)\,dy+(1-q)\pi_B^L\int_{-\infty}^{\hat{y}(t,q)}e^{-\frac{1}{2}t+\sqrt{t}y}\phi(y)\,dy, \tag{35}$$

where $\sqrt{t}\hat{y}(t,q)=t/2+\log\left(\left(\frac{M-m}{\pi_A^L-\pi_B^L}-1\right)\frac{q}{1-q}\right)$ satisfies:

$$\pi_A^L+\frac{m}{\left(\frac{1}{q}-1\right)e^{-\frac{1}{2}t+\sqrt{t}\hat{y}(t,q)}+1}=\pi_B^L+\frac{M}{\left(\frac{1}{q}-1\right)e^{-\frac{1}{2}t+\sqrt{t}\hat{y}(t,q)}+1}.$$

The second and the last integrands in (35) can be simplified using

$$e^{-\frac{1}{2}t+\sqrt{t}y}\phi(y)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}t+\sqrt{t}y-\frac{y^2}{2}}=\frac{1}{\sqrt{2\pi}}e^{-\frac{(y-\sqrt{t})^2}{2}},$$

i.e. the pdf of a normal variable with mean $\sqrt{t}$ and unit variance. With (35), we get

$$u(t,q)=q\left(\pi_A^L+m\right)\int_{\hat{y}(t,q)}^{\infty}\phi(y)\,dy+(1-q)\pi_A^L\int_{\hat{y}(t,q)-\sqrt{t}}^{\infty}\phi(y)\,dy$$

$$+q\left(\pi_B^L+M\right)\Phi\left(\hat{y}(t,q)\right)+(1-q)\pi_B^L\Phi\left(\hat{y}(t,q)-\sqrt{t}\right).$$

Symmetry $\phi(y)=\phi(-y)$ and all parametric definitions yield

$$u(t,q)=q\pi_A^H\Phi\left(-\frac{1}{2}\sqrt{t}+\frac{1}{\sqrt{t}}L(\hat{q},q)\right)+(1-q)\pi_A^L\Phi\left(\frac{1}{2}\sqrt{t}+\frac{1}{\sqrt{t}}L(\hat{q},q)\right)$$

$$+q\pi_B^H\Phi\left(\frac{1}{2}\sqrt{t}-\frac{1}{\sqrt{t}}L(\hat{q},q)\right)+(1-q)\pi_B^L\Phi\left(-\frac{1}{2}\sqrt{t}-\frac{1}{\sqrt{t}}L(\hat{q},q)\right).$$

Using $\Phi(y)=1-\Phi(-y)$, we get $FIG(t,q)$; $v(t,q)=u(t,q)-u(q)$ gives (13).

A.4. Marginal value: Proof of Theorem 2

Differentiating Theorem 1 in t, and denote $L=L(\hat{q},q)$. Then $v_t(t,q)$ equals

$$q\pi_A^H\phi\left(-\frac{\sqrt{t}}{2}+\frac{L}{\sqrt{t}}\right)\left(-\frac{1}{4\sqrt{t}}-\frac{L}{2t^{3/2}}\right)+(1-q)\pi_A^L\phi\left(\frac{\sqrt{t}}{2}+\frac{L}{\sqrt{t}}\right)\left(\frac{1}{4\sqrt{t}}-\frac{L}{2t^{3/2}}\right)$$

$$+q\pi_B^H\phi\left(\frac{\sqrt{t}}{2}-\frac{L}{\sqrt{t}}\right)\left(\frac{1}{4\sqrt{t}}+\frac{L}{2t^{3/2}}\right)+(1-q)\pi_B^L\phi\left(-\frac{\sqrt{t}}{2}-\frac{L}{\sqrt{t}}\right)\left(-\frac{1}{4\sqrt{t}}+\frac{L}{2t^{3/2}}\right)$$

$$=\phi\left(-\frac{\sqrt{t}}{2}+\frac{L}{\sqrt{t}}\right)\left[(\pi_B^H-\pi_A^H)q\left(\frac{1}{4\sqrt{t}}+\frac{L}{2t^{3/2}}\right)+(\pi_A^L-\pi_B^L)\frac{q(1-\hat{q})}{\hat{q}}\left(\frac{1}{4\sqrt{t}}-\frac{L}{2t^{3/2}}\right)\right]$$

$$=\frac{(M-m)q(1-\hat{q})}{2\sqrt{t}}\phi\left(-\frac{\sqrt{t}}{2}+\frac{L}{\sqrt{t}}\right),$$

where the second equality owes to $\phi(-x) = \phi(x)$ and $\phi\left(\frac{\sqrt{t}}{2} + \frac{L}{\sqrt{t}}\right) = \phi\left(-\frac{\sqrt{t}}{2} + \frac{L}{\sqrt{t}}\right)e^{-L}$, and the last to $M - m = (\pi_B^H - \pi_A^H) + (\pi_A^L - \pi_B^L)$ and $\hat{q} = (\pi_A^L - \pi_B^L)/(M - m)$.

Finally (15) yields (14) by taking time derivatives and by using Lemma 5.

A.5. Slopes at zero: Proof of Corollary 1

Claim 1. *For any $n > 1$ we have*:

$$\frac{\partial^n v(t,q)}{\partial t^n} = v_t(t,q)\left[\frac{A_{2(n-1)}(q)}{t^{2(n-1)}} + \cdots + \frac{A_1(q)}{t} + A_0(q)\right],$$

where $A_{2(n-1)}(q), \ldots, A_0(q)$ are bounded functions of q.

Proof. From (9), since $v_{tt}(t,q) = v_t(t,q)\left[\frac{A_2(q)}{t^2} + \frac{A_1(q)}{t} + A_0(q)\right]$, each differentiation produces a polynomial in $1/t$ whose highest power is greater by two. □

The first equality of Corollary 1 owes to (15) because $\phi\left(-\frac{1}{2}\sqrt{t} + \frac{1}{\sqrt{t}}L(\hat{q},q)\right) > 0$ for all $(t,q) \in (0,\infty) \times (0,1)$. The second, third, and fourth equalities owe to (8), (9), and (15), by taking the limit $t \to \infty$ or $t \downarrow 0$. [Indeed, $\lim_{t\downarrow 0} \xi(t,q,\hat{q}) = \lim_{s\downarrow 0} \frac{1}{\sqrt{s}} e^{-\frac{1}{s}} e^{-s} = 0$ if $q \neq \hat{q}$.] From Lemmas 1, 6, and Theorem 2, we get

$$\frac{\partial^n v(t,q)}{\partial t^n} = \sqrt{\frac{q(1-q)}{\hat{q}^3(1-\hat{q})^3 2\pi}} \frac{1}{e^{\frac{1}{8}t + \frac{1}{2t}L^2(\hat{q},q)}} \frac{1}{\sqrt{t}} \left[\frac{A_{2(n-1)}(q)}{t^{2(n-1)}} + \cdots + \frac{A_1(q)}{t} + A_0(q)\right] = 0.$$

This gives the fifth and sixth equality of Corollary 1. □

A.6. Information lumpiness: Proof of Theorem 5

Because $T_{\mathrm{CH}}(q) \geqslant T_{\mathrm{FL}}(q)$, the last integral in the proof is greater than

$$\int_0^{T_{\mathrm{FL}}} \frac{e^{-\frac{T_{\mathrm{FL}}}{8} - L^2/2s}}{2\sqrt{s}}\,ds = \begin{cases} e^{-\frac{T_{\mathrm{FL}}}{8}}\left[\sqrt{T_{\mathrm{FL}}}e^{-\frac{L^2}{2T_{\mathrm{FL}}}} - \sqrt{2\pi}L\left(1 - \Phi\left(\frac{L}{\sqrt{T_{\mathrm{FL}}}}\right)\right)\right] & \forall \hat{q} \geqslant q, \\ e^{-\frac{T_{\mathrm{FL}}}{8}}\left[\sqrt{T_{\mathrm{FL}}}e^{-\frac{L^2}{2T_{\mathrm{FL}}}} + \sqrt{2\pi}L\Phi\left(\frac{L}{\sqrt{T_{\mathrm{FL}}}}\right)\right] & \forall q \geqslant \hat{q}. \end{cases} \tag{36}$$

Eq. (36) gives that $v(T_{\mathrm{CH}}(q),q)/[\hat{q}(1-\hat{q})(M-m)]$ exceeds

$$j(q,\hat{q}) = \begin{cases} \sqrt{\frac{q(1-q)}{2\pi\hat{q}(1-\hat{q})}} e^{-\frac{1}{8}T_{\mathrm{FL}}}\left[\sqrt{T_{\mathrm{FL}}}e^{-\frac{L^2}{2T_{\mathrm{FL}}}} - \sqrt{2\pi}L\left(1 - \Phi\left(\frac{L}{\sqrt{T_{\mathrm{FL}}}}\right)\right)\right] & \forall \hat{q} \geqslant q, \\ \sqrt{\frac{q(1-q)}{2\pi\hat{q}(1-\hat{q})}} e^{-\frac{1}{8}T_{\mathrm{FL}}}\left[\sqrt{T_{\mathrm{FL}}}e^{-\frac{L^2}{2T_{\mathrm{FL}}}} + \sqrt{2\pi}L\Phi\left(\frac{L}{\sqrt{T_{\mathrm{FL}}}}\right)\right] & \forall q \geqslant \hat{q}. \end{cases}$$

One can verify that $J(\hat{q}) = \int_0^1 j(q, \hat{q})\, dq$ is convex on (0, 1), and is minimized at $\hat{q} = 0.5$. Now, $j(\cdot, 0.5)$ is a double-hump shape: it is concave on (0, 0.5) and (0.5, 1), and $\lim_{q\downarrow 0} j(q, 0.5) = \lim_{q\to 0.5} j(q, 0.5) = \lim_{q\uparrow 1} j(q, 0.5) = 0$ because $L(0.5, q) = 0$ and $T_{\mathrm{FL}} = 0$ if $\hat{q} = q = 0.5$, while the limits at $q = 0$ and 1 require l'Hopital's rule. Further, $j(\cdot, 0.5)$ satisfies $j(r, 0.5) = j(1 - r, 0.5)$ where $0 < r < 0.5$. So we can inscribe between the horizontal $j = 0$ axis and the $j(q, 0.5)$ curve two equally tall triangles, whose area is a lower bound on $J(0.5)$, namely, $J(0.5) > (0.5) \max_{q\in(0,0.5)} j(q, 0.5)$. Finally, $\max_{q\in(0,0.5)} j(q, 0.5) = j(0.25, 0.5) > 0.05$.

A.7. Elasticity of demand: Proof of Theorem 8

Proof of Part (a). Now we show that there exist two points where $T_E(q) = T_{\mathrm{CH}}(q)$—one when $q < \hat{q}$ and one when $q > \hat{q}$. Define the *gross surplus function* γ as follows:

$$\gamma(q) \equiv v(T_E(q), q) - v_t(T_E(q), q)T_E(q). \tag{37}$$

It suffices to show that $\gamma(q) > 0$ iff $q \in (\acute{q}, \grave{q})$, where $\hat{q} \in (\acute{q}, \grave{q})$. We prove that $\gamma(q) > 0$ for $q \in (\acute{q}, \hat{q})$; the case $q \in (\hat{q}, \grave{q})$ is similar. Differentiating (8) yields:

Claim 2 (*q-Derivatives*). *The transition pdf* $q(t)$ *obeys, for* $0 < q, r < 1$ *and* $t > 0$:

$$\xi_q(t, q, r) = \xi(t, q, r)\left[\frac{1 - 2q}{2q(1 - q)} + \frac{L(r, q)}{tq(1 - q)}\right],$$

$$\xi_{qq}(t, q, r) = \xi(t, q, r)\frac{1}{q^2(1 - q)^2}\left[-\frac{1}{4} + \frac{L^2(r, q)}{t^2} - \frac{1}{t}\right]. \tag{38}$$

Claim 3. *We have* $\gamma(0+) = 0$, $\gamma_q(0+) = -\infty$, *and* $\gamma(\hat{q}) > 0$.

Proof. First we calculate $\gamma(0+)$. Since $\lim_{q\downarrow 0} L(\hat{q}, q) = \lim_{q\downarrow 0} \log\left(\frac{\hat{q}(1-q)}{q(1-\hat{q})}\right) = \infty$, we get $\lim_{q\downarrow 0} T_E(q) = \lim_{q\downarrow 0} L^2(\hat{q}, q) = \infty$ and using (8) and (15), we get

$$\lim_{q\downarrow 0} v_t\,(T_E(q), q)\, T_E(q) = \lim_{q\downarrow 0} \frac{\sqrt{qT_E(q)}}{e^{\frac{T_E(q)}{8} + \frac{1}{2T_E(q)}L^2(\hat{q},q)}} = 0.$$

From (13), the first term of $v(T_E(q), q)$ satisfies

$$\lim_{q\downarrow 0} q(\pi_B^H - \pi_A^H)\Phi\left(\frac{\sqrt{T_E(q)}}{2} - \frac{L(\hat{q}, q)}{\sqrt{T_E(q)}}\right) = 0,$$

because $\Phi(r) \leqslant 1$ for all r. For the second term of $v(T_E(q), q)$, we have

$$\lim_{q\downarrow 0}(1 - q)(\pi_A^L - \pi_A^L)\Phi\left(-\frac{\sqrt{T_E(q)}}{2} - \frac{L(\hat{q}, q)}{\sqrt{T_E(q)}}\right) = 0$$

since $\lim_{q\downarrow 0} L(\hat{q}, q)/\sqrt{T_E(q)} = \lim_{q\downarrow 0} \sqrt{T_E(q)} = \infty$. Thus, $\lim_{q\downarrow 0} \gamma(q) = 0$.

Next we solve $\lim_{r\downarrow 0}\gamma_q(r)$, where

$$\gamma_q(q)=v_q(T_E(q),q)-v_{tt}(T_E(q),q)T_E(q)\frac{\partial}{\partial q}T_E(q)-v_{tq}(T_E(q),q)T_E(q). \tag{39}$$

We analyze the different terms separately. By Theorem 3, we have $v_{tt}(T_E(q),q)<0$. Differentiating (25) gives

$$\frac{\partial}{\partial q}T_E(q)=-\frac{2}{q(1-q)}\frac{L(\hat{q},q)}{\sqrt{1+L^2(\hat{q},q)}},$$

and so $\lim_{r\downarrow 0}\frac{\partial}{\partial q}T_E(r)=-\infty$. Because $\lim_{q\downarrow 0}T_E(q)=\infty$, the second term in (39) satisfies $\lim_{q\downarrow 0}v_{tt}(T_E(q),q)T_E(q)\frac{\partial}{\partial q}T_E(q)=\infty$. For the third term of (39):

$$\lim_{r\downarrow 0}v_{tq}(T_E(r),r)=\lim_{q\downarrow 0}\xi(T_E(q),q,\hat{q})\left[\frac{1-2q}{2q(1-q)}+\frac{L(\hat{q},q)}{T_E(q)q(1-q)}\right]=\infty,$$

because $\lim_{q\downarrow 0}\frac{q}{T_E(q)}=0$. Since $v_q(T_E,q)$ is bounded, we get $\gamma_q(0+)=-\infty$.

Finally, compute the value of $\gamma(\hat{q})$ by using (37). We again divide the expression into different parts. Since $L(\hat{q},\hat{q})=0$, we have $T_E(\hat{q})=4$, and so from (15), we get

$$v_t(T_E(\hat{q}),\hat{q})T_E(\hat{q})=\hat{q}(1-\hat{q})(M-m)\sqrt{\frac{1}{2\pi}}e^{-1/2}. \tag{40}$$

Since $T_E(\hat{q})=4$, Eq. (13) gives that $v(T_E(\hat{q}),\hat{q})$ equals

$$\hat{q}(1-\hat{q})(M-m)\,(\Phi(1)-\Phi(-1))=\hat{q}(1-\hat{q})(M-m)\sqrt{\frac{1}{2\pi}}\int_{-1}^{1}e^{-\frac{x^2}{2}}\,dx. \tag{41}$$

By subtracting (40) from (41), we get $\gamma(\hat{q})>0$, since $\int_{-1}^{1}e^{-\frac{x^2}{2}}\,dx>e^{-1/2}$. □

Claim 4. *$\gamma(\cdot):(0,\hat{q})\to\mathbb{R}$ is convex.*

Proof. Twice differentiating (37), the second derivative $-\gamma_{qq}(q)$ equals

$$\begin{aligned}&v_{ttt}(T_E(q),q)T_E(q)\left(\frac{\partial}{\partial q}T_E(q)\right)^2+2v_{ttq}(T_E(q),q)T_E(q)\frac{\partial}{\partial q}T_E(q)\\&\quad+v_{tt}(T_E(q),q)\left(\frac{\partial}{\partial q}T_E(q)\right)^2+v_{tt}(T_E(q),q)T_E(q)\frac{\partial^2}{\partial q^2}T_E(q)\\&\quad+v_{tqq}(T_E(q),q)T_E(q)-v_{qq}(T_E,q).\end{aligned}$$

Substitute from (17), (18), and the formulas below from Claim 2 and Theorem 2:

$$v_{ttq}(t,q)=v_t(t,q)\left[\left(-\frac{1}{8}+\frac{L^2}{2t^2}-\frac{1}{2t}\right)\left(\frac{1-2q}{2q(1-q)}+\frac{L}{tq(1-q)}\right)-\frac{L}{t^2q(1-q)}\right],$$

$$v_{tq}(t,q)=v_t(t,q)\left[\frac{1-2q}{2q(1-q)}+\frac{L}{tq(1-q)}\right],$$

$$v_{tqq}(t,q)/v_{tt}(t,q) = v_{qq}(t,q)/v_t(t,q) = \frac{2}{q^2(1-q)^2}$$

$$\Longrightarrow \gamma_{qq}(q) = v_t(T_E(q),q)\frac{\left(3+L^2\right)+\left(L^4+4L^2(2+S)+8(1+S)\right)}{(1-q)^2q^2\left(1+L^2\right)^{\frac{3}{2}}S^3} > 0,$$

where $S = 1+\sqrt{1+L^2}$. □

From Claim 3 we get $\gamma(\hat{q}) > 0$ and $\gamma(\varepsilon) < 0$ for $\varepsilon > 0$. Since $\gamma(\cdot)$ is continuous, there exists $\acute{q} \in (0,\hat{q})$ with $\gamma(\acute{q}) = 0$. Claim 4 gives uniqueness. So $\gamma(q) > 0$ for $q \in (\acute{q},\hat{q})$. The existence/uniqueness proof for $\grave{q} \in (\hat{q},1)$ is symmetric. □

Proof of Part (b). Note that when $c \leqslant c_{\mathrm{CH}}(q)$ then $\tau(c,q) > 0$. Let us denote elasticity $E(c) = -c\tau_c(c,q)/\tau(c,q)$. Clearly, $E'(c) > 0$ iff

$$c\tau_c^2(c,q) - [\tau_c(c,q)+\tau_{cc}(c,q)c]\,\tau(c,q) > 0. \tag{42}$$

Differentiating $v_t(\tau(c,q),q) = c$ yields $\tau_c(c,q) = 1/v_{tt}$ and $\tau_{cc}(c,q) = -v_{ttt}(\tau,q)/v_{tt}^3(\tau,q)$. Hence, if we substitute from (17) and (18) for v_{tt}/v_t and v_{ttt}/v_{tt}, we get

$$\begin{aligned}
c\tau_c^2 - \tau(\tau_c+\tau_{cc}c) &= \frac{v_t}{v_{tt}^2} - \tau\left(\frac{1}{v_{tt}} - \frac{v_{ttt}v_t}{v_{tt}^3}\right)\\
&= -\frac{\tau}{v_{tt}}\left(1 - \frac{v_t^2}{v_{tt}^2}\left(\left(\frac{v_{tt}}{v_t}\right)^2 - \frac{L^2}{\tau^3} + \frac{1}{2\tau^2}\right) - \frac{v_t}{\tau v_{tt}}\right)\\
&= -\frac{1}{v_{tt}(v_{tt}/v_t)^2\tau^2}\left(L^2 - \frac{1}{2}\tau - \frac{v_{tt}}{v_t}\tau^2\right)\\
&= -\frac{1}{v_{tt}(v_{tt}/v_t)^2\tau^2}\left(\frac{L^2}{2} + \frac{\tau^2}{8}\right)
\end{aligned}$$

which is positive because $v_{tt}(\tau,q) < 0$ when $\tau(c,q) > T_{\mathrm{CH}}(q)$. Hence, $E(c)$ is rising in the cost c, and thus falling in the quantity τ. □

A.8. Large demand: inverse demand curve of Theorem 9

Claim 5. *Assume that* $\varepsilon(x) > 0$ *is an increasing* C^1 *function of* x, *with* $\varepsilon(x)/x \to 0$, *and* $\varepsilon'(x) = \varsigma/x + O(1/x^2)$. *Then the map* $\psi(x) = x + \varepsilon(x)$ *has inverse* $\beta(x) = x - \delta(x)$ *where* $\delta(x) = \varepsilon(x)(1 - \varsigma/x + O(1/x^2))$. *Furthermore*, $\delta(x) < \varepsilon(x)$ *for all* x.

Notice that $\varepsilon(x) = \frac{1}{2}\log x + B/x$ obeys the required conditions with $\varsigma = \frac{1}{2}$.

Proof of Claim. Let $\beta(x) = x - \delta(x)$, for $\delta(x) > 0$—whose sign is clear, because β and ψ are reflections of each other about the diagonal. Also, since $\psi(x) \to \infty$, so must $\beta(x) \to \infty$, by reflection. By the inverse property, $\psi(x-\delta(x)) \equiv x - \delta(x) + \varepsilon(x-\delta(x)) \equiv x$. Since $x \mapsto \varepsilon(x)$ is increasing, $\delta(x) = \varepsilon(x-\delta(x)) < \varepsilon(x)$.

Taking a first order Taylor series of ε about x yields $\delta(x) = \varepsilon(x) - \delta(x)\varepsilon'(\hat{x}) < \varepsilon(x)$ for some intermediate $\hat{x} \in [x - \delta(x), x]$. Hence, $\hat{x}/x \geqslant 1 - \delta(x)/x \geqslant 1 - \varepsilon(x)/x \to 1$.

$$\delta(x) = \frac{\varepsilon(x)}{1 + \varepsilon'(\hat{x})} = \varepsilon(x)(1 - \varsigma/\hat{x} + O(1/\hat{x}^2)) = \varepsilon(x)(1 - \varsigma/\hat{x} + O(1/x^2)).$$

A.9. Convergent belief processes: Proof of Theorem 10

A.9.1. Preliminary lemmas

Recall that $\lambda(Z|\Delta) = dG(Z|L, \Delta)/dG(Z|H, \Delta)$. Of course, since we are proving Theorem 10, we maintain (29).

Lemma 9. *Mean and variance obey*: $Var\left[\lambda(Z|\Delta)|H\right] = E\left[\lambda(Z|\Delta)|L\right] - 1$.

Proof. Since $E\left[\lambda(Z|\Delta)|H\right] = 1$, the variance in H equals

$$\int [1 - \lambda(z|\Delta)]^2 \, dG(z|H, \Delta) = \int \left[1 + \lambda^2(z|\Delta) - 2\lambda(z|\Delta)\right] dG(z|H, \Delta)$$
$$= \int \left(\frac{dG(z|L, \Delta)}{dG(z|H, \Delta)}\right)^2 dG(z|H, \Delta) - 1. \qquad \square$$

Lemma 10. $E[\lambda(Z|\Delta)|L] = 1 + \Delta + o(\Delta)$ *and* $Var[\lambda(Z|\Delta)|L] = 1 + o(\Delta)$.

Proof. The first equality follows from Lemma 9 and (29), because

$$E[\lambda(Z|\Delta)|L] = 1 + Var[\lambda(Z|\Delta)|H] = 1 + \Delta + o(\Delta)$$
$$\Rightarrow Var\left[\lambda(Z|\Delta)|L\right] = \int \left[\lambda(z|\Delta) - 1 - \Delta - o(\Delta)\right]^2 dG(|\Delta, L)$$
$$= \int \left[1 - \lambda(z|\Delta)\right]^2 dG(z|L, \Delta) + o(\Delta)$$

which equals $\Delta + o(\Delta)$ by (29). $\square$

Lemma 11 (*A derived process*). *Let* W_θ *be a Wiener process in state* θ. *The process* $S_t \equiv \sum_{i=1}^{\lfloor t/\Delta \rfloor} (1 - \lambda(Z_i|\Delta))$ *converges weakly to* $W_H(t)$ *in state* H *as* $\Delta \downarrow 0$, *and to* $-t + W_L(t)$ *in state* L. *Therefore*, $dW_H = -dt + dW_L$.

Proof. By [3] (Theorem 8.7.1), three conditions give weak convergence: the limiting discrete belief process almost surely has continuous sample paths, and the first two moments of $S_t - S_{t-\Delta}$ converge to their continuous time analogs.

Step 1: Limit continuity. For each $\varepsilon > 0$ and $\theta \in \{H, L\}$, we have

$$P_\theta^\Delta \left(|1 - \lambda(Z|\Delta)| \geqslant \varepsilon\right) \leqslant \varepsilon^{-1} \int |1 - \lambda(z|\Delta)| \, dG(z|\theta, \Delta)$$

where P_θ^Δ is the probability measure for $G(\cdot|\theta, \Delta)$. By (29), this vanishes in Δ.

Steps 2 *and* 3: Convergent first and second moments. We have

$$\lim_{\Delta\downarrow 0} \tfrac{1}{\Delta} E\left[S_t - S_{t-\Delta}|H\right] = 0 = E[dW_H|H]/dt,$$

$$\lim_{\Delta\downarrow 0} \tfrac{1}{\Delta} E\left[(S_t - S_{t-\Delta})^2|H\right] = 1 = E[(dW_H)^2|H]/dt$$

by the martingale property and (29). Likewise, from Lemma 10 and (29) we get

$$\lim_{\Delta\downarrow 0} \tfrac{1}{\Delta} E\left[S_t - S_{t-\Delta}|L\right] = -1 = E[-dt + dW_L|L]/dt,$$

$$\lim_{\Delta\downarrow 0} \tfrac{1}{\Delta} E\left[(S_t - S_{t-\Delta} + \Delta + o(\Delta))^2|L\right] = 1 = E[(dW_L)^2|L]/dt. \qquad \square$$

A.9.2. Proof of Theorem 10

We first derive the process of *reciprocal beliefs* $1/q(t)$, and from it deduce the belief process. Bayes rule (28) yields a nice expression:

$$\frac{1}{q^\Delta(n\Delta)} - \frac{1}{q^\Delta((n-1)\Delta)} = \left(1 - \frac{1}{q^\Delta((n-1)\Delta)}\right)(1 - \lambda(Z_n|\Delta)). \tag{43}$$

Step 1: $\alpha(q) = 1/q$ obeys $d\alpha(q(t)) = [1 - \alpha(q(t))]\, dW_H(t)$ in State H. By (43):

$$\alpha\left(q^\Delta(t)\right) - \alpha\left(q^\Delta(0)\right) = \sum_{i=1}^{\lfloor t/\Delta \rfloor} \left[1 - \alpha\left(q^\Delta((i-1)\Delta)\right)\right]\left[1 - \lambda(Z_i|\Delta)\right]. \tag{44}$$

We wish to treat these as partial sums for an Ito integral. To do so, we must ensure that the limits are well-defined. Hence, we define stopping times:

$$t_n = \sup\left\{t : E\left[\lim_{\Delta\downarrow 0} \sum_{i=1}^{\lfloor t/\Delta \rfloor} \left[1 - \alpha\left(q^\Delta((i-1)\Delta)\right)\right]^2 \Delta \,\middle|\, H\right] \leqslant n\right\}.$$

By the construction of the Ito integral, we take limits in (44) using Lemma 11:

$$\lim_{\Delta\downarrow 0}\left\{\alpha\left(q^\Delta(t \wedge t_n)\right) - \alpha\left(q^\Delta(0)\right)\right\} = \int_0^{t\wedge t_n} [1 - \alpha(q(s))]\, dW_H(s), \tag{45}$$

where the Ito integrals are well defined by the definition of t_n.

If $t_n \to \infty$, then we have the stronger assertion than (45), that for $t < \infty$:

$$\lim_{\Delta\downarrow 0}\left\{\alpha(q^\Delta(t)) - \alpha(q^\Delta(0))\right\} = \int_0^t [1 - \alpha(q(s))]\, dW_H(s). \tag{46}$$

Let $t \leqslant t_n$, where (46) holds. Then $Y(t) = 1 - \alpha(q(t))$ obeys $dY(t) = -d\alpha(q(t)) = -Y(t)\, dW_H(t)$, i.e. $Y(t_n) = Y(0)\exp(-t_n/2 - W_H(t_n))$. Thus

$$E\left[\int_0^{t_n} Y^2(s)\, ds \,\middle|\, H\right] = Y^2(0)\int_0^{t_n} e^s\, ds = Y^2(0)\left(e^{t_n} - 1\right) < \infty$$

if t_n is boundedly finite. This precludes $n \to \infty$, contradiction. So $t_n \to \infty$.

Step 2: The process for $q(t) = 1/\alpha(q(t))$. By Ito's Lemma, Step 1, and Lemma 11, we respectively have for all $t \geqslant 0$,

$$\begin{aligned} dq(t) &= -\frac{1}{\alpha^2(q)} d\alpha(q) + \frac{1}{\alpha^3(q)} (d\alpha(q))^2 \\ &= q(t)(1 - q(t)) \, [dW_H(t) + (1 - q(t)) \, dt] \\ &= q(t)(1 - q(t)) \, [dW_L(t) - q(t) \, dt] \, . \end{aligned}$$

Define the unconditional Wiener process as $dW = q \, dW_H + (1 - q) \, dW_L$. Then

$$dq = q(1 - q) \, (q[dW_H + (1 - q) \, dt] + (1 - q)[dW_L - q \, dt]) = q(1 - q) \, dW$$

yields the unconditional belief process, as desired.

A.10. Approximate value functions: Proof of Theorem 11

For every $y \geqslant 0$, let

$$\Omega^\Delta(y\Delta|c) \equiv \Pi^\Delta \left(\lfloor y + 1 \rfloor \Delta|c\right) [1 - (y - \lfloor y \rfloor)]^{\frac{1}{\Delta}} + \Pi^\Delta \left(\lfloor y \rfloor \Delta|c\right) (y - \lfloor y \rfloor)^{\frac{1}{\Delta}} \, .$$

This is an average of the maximand $\Pi^\Delta(y\Delta|c)$ of (33) at the integers adjacent to y. Also, $\Omega^\Delta(y\Delta|c)$ is continuous in $y \geqslant 0$, $c \geqslant 0$, and $\Delta > 0$. The latter holds because $c\lfloor y \rfloor \Delta$ is continuous in Δ, and $v^\Delta(\lfloor y \rfloor \Delta)$ is continuous in Δ for given $\lfloor y \rfloor$. Thus, $\Pi^\Delta(\lfloor y \rfloor \Delta|c) = v^\Delta \left(\lfloor y \rfloor \Delta\right) - c \lfloor y \rfloor \Delta$ is continuous in Δ, and so therefore is $\Omega^\Delta(y\Delta|c)$.

Next, $\Omega^\Delta(y\Delta|c)$ coincides with the discrete maximand at $\Delta, 2\Delta, \ldots$. Also, $\Omega^\Delta(y\Delta|c)$ is an average of $\Pi^\Delta \left(\lfloor y \rfloor \Delta|c\right)$ and $\Pi^\Delta \left(\lfloor y + 1 \rfloor \Delta|c\right)$, and so of $\Omega^\Delta \left(\lfloor y \rfloor \Delta|c\right)$ and $\Omega^\Delta \left(\lfloor y + 1 \rfloor \Delta|c\right)$. Then $\Omega^\Delta(y\Delta|c) \leqslant \max \left\langle \Omega^\Delta \left(\lfloor y \rfloor \Delta|c\right), \Omega^\Delta \left(\lfloor y + 1 \rfloor \Delta|c\right) \right\rangle$ with strict inequality iff y is not an integer and $\Pi^\Delta \left(\lfloor y + 1 \rfloor \Delta|c\right) \neq \Pi^\Delta \left(\lfloor y \rfloor \Delta|c\right)$. As we can improve weakly over any y by choosing either $\lfloor y \rfloor$ or $\lfloor y + 1 \rfloor$, the correspondence $M^\Delta(c) \equiv \arg\max_{y \geqslant 0} \Omega^\Delta(y\Delta|c)$ contains a non-negative integer. Altogether

$$\max_{y \geqslant 0} \Omega^\Delta(y\Delta|c) = \max_{n=0,1,2,\ldots} \Pi^\Delta(n\Delta|c). \tag{47}$$

Finally, $0 \leqslant y - \lfloor y \rfloor \leqslant 1$, and so $(y - \lfloor y \rfloor)^{\frac{1}{\Delta}}$ is a positive function of Δ that vanishes with Δ but remains continuous in y for every $\Delta > 0$. So for every given $t > 0$,

$$\lim_{\Delta \downarrow 0} \Omega^\Delta(t|c) = \lim_{\Delta \downarrow 0} v^\Delta(t) - ct = v^0(t) - ct = \Pi(t|c).$$

We are now ready to use the auxiliary problem of maximizing $\Omega^\Delta(y\Delta|c)$ over $y \geqslant 0$. Again, $\Omega^\Delta \,(0|c) = 0$ and $\lim_{y \to \infty} v^\Delta(\lfloor y \rfloor \Delta) \leqslant \max_{\theta,a} \pi_a^\theta < \infty$. We can thus restrict the choice of y to a compact interval $[0, \bar{y}(\Delta)]$, where $\bar{y}(\Delta)$ is the continuous function defined by the largest solution m to $v^\Delta \,(m\Delta) = cm\Delta$. We can then rewrite

$$M^\Delta(c) = \arg \max_{y \in [0, \bar{y}(\Delta)]} \Omega^\Delta(y\Delta|c).$$

Demand maximizes a function $\Omega^\Delta(y\Delta|c)$ continuous in y, c, Δ over a compact-valued and continuous correspondence $[0, \bar{y}(\Delta)]$. A non-integer y belongs to $M^\Delta(c)$ iff $\lfloor y \rfloor$ and $\lfloor y + 1 \rfloor$ do.

This is a non-generic event in c; generically, $M^{\Delta}(c)$ contains either one integer, or non-consecutive integers. Thus, $M^{\Delta}(c)$ contains only integers a.e. in c, Δ. The maximized Ω^{Δ} coincides with the discrete maximand $\Pi^{\Delta}(n\Delta|c)$:

$$\max_{y\in[0,\bar{y}(\Delta)]} \Omega^{\Delta}(y\Delta|c) = \max_{y\in[0,\bar{y}(\Delta)]} \Omega^{\Delta}(\lfloor y \rfloor \Delta|c) = \max_{n=0,1,2,\ldots} \Pi^{\Delta}(n\Delta|c)$$

we have $M^{\Delta}(c) = N^{\Delta}(c)$ a.e. in parameter space. Write the first maximization as

$$\max_{t\in[0,\bar{y}(\Delta)\Delta]} \Omega^{\Delta}(t|c)$$

the maximization over a compact-valued, continuous correspondence $[0, \bar{y}(\Delta)\Delta]$ of a function $\Omega^{\Delta}(t|c)$ continuous in t, c, Δ. By the Theorem of the Maximum, the correspondence $T^{\Delta}(c) = M^{\Delta}(c)/\Delta$ is u.h.c. in Δ and c. As τ is single-valued at all but one c, as $\Delta \downarrow 0$, all selections $\tau^{\Delta}(c, q) \in T^{\Delta}(c)$ converge to the unique maximizer $\tau(c, q)$ of the continuous time problem $\Pi(t|c) = \lim_{\Delta\downarrow 0} \Omega^{\Delta}(t|c)$: namely, $\lim_{\Delta\downarrow 0} \left|\tau^{\Delta}(c, q) - \tau(c, q)\right| = 0$ a.e. in parameter space. Let $y^{\Delta}(c) \equiv \tau^{\Delta}(c, q)/\Delta \in M^{\Delta}(c)$. This selection must be integer-valued and a maximizer of $\Pi^{\Delta}(n\Delta|c)$ a.e. in parameter space. So $y^{\Delta}(c) = n^{\Delta}(c)$ for some optimal discrete sample size $n^{\Delta}(c) = \tau^{\Delta}(c, q)/\Delta \in N^{\Delta}(c, q)$. Then a.e. in parameter space,

$$0 = \lim_{\Delta\downarrow 0} \left|\tau^{\Delta}(c) - \tau(c, q)\right| = \lim_{\Delta\downarrow 0} \left|y^{\Delta}(c)\Delta - \tau(c, q)\right| = \lim_{\Delta\downarrow 0} \left|n^{\Delta}(c)\Delta - \tau(c, q)\right|.$$

References

[1] F. Black, M.S. Scholes, The pricing of options and corporate liabilities, J. Polit. Economy 81 (1973) 637–654.
[2] H. Chade, E. Schlee, Another look at the Radner-Stiglitz nonconcavity in the value of information, J. Econ. Theory 107 (2002) 421–452.
[3] R. Durrett, Stochastic Calculus: a Practical Introduction, CRC Press, Boca Raton, FL, 1996.
[4] J.M. Harrison, D. Kreps, Martingales and arbitrage in multiperiod securities markets, J. Econ. Theory 20 (1979) 381–408.
[5] J.M. Harrison, S.R. Pliska, Martingales and stochastic integrals in the theory of continuous trading, Stochast. Process. Appl. 11 (1981) 215–260.
[6] I. Karatzas, S.E. Shreve, Brownian Motion and Stochastic Calculus, second ed., Springer, Berlin, 1991.
[7] R. Kihlstrom, A Bayesian model of demand for information about product quality, Int. Econ. Rev. 15 (1974) 99–118.
[8] R.S. Liptser, A.N. Shiryayev, Statistics of Random Processes, vol. I, second ed., Springer, Berlin, 2001.
[9] G. Moscarini, L. Smith, The law of large demand for information, Econometrica 70 (6) (2001) 2351–2366.
[10] G. Moscarini, L. Smith, The optimal level of experimentation, Econometrica 69 (2001) 1629–1644.
[11] B. Øksendal, Stochastic Differential Equations: An Introduction with Applications, vol. I, fifth ed., Springer, Berlin, 1998.
[12] R. Radner, J. Stiglitz, A nonconcavity in the value of information, in: M. Boyer, R. Kihlstrom (Eds.), Bayesian Models in Economic Theory, North-Holland, Amsterdam, 1984, pp. 33–52.
[13] C.E. Shannon, A mathematical theory of communication, Bell System Tech. J. 27 (1948) 379–423, 623–656.