Monopoly Pricing under Mutual and Directed Information Constraints

Stefan Behringer

Value of Information Group, Universität Bielefeld, Germany

Abstract

We study monopoly pricing under demand and cost uncertainty using information-theoretic constraints. Rather than specifying additive or multiplicative noise directly, we take mutual and directed information between market fundamentals and observable demand as primitive. Mapping the additive-multiplicative noise model of Behringer (2021) into an information capacity shows that its non-monotone value of information is an artefact of the noise-variance parameterisation: recast in terms of the Value of Information (VoI) in the Stratonovich/Shannon sense (Behringer, 2026), with Shannon information as the argument, the pathology disappears. Directed information lets us treat endogenous pricing and feedback explicitly, and the resulting VoI is monotone and concave in Shannon information in both static and dynamic environments. Building on this foundation, we cast the firm as a rationally inattentive monopolist that acquires and allocates scarce attention across two channels, demand and cost. The optimal split follows a reverse water-filling rule governed by the ratio of demand to cost uncertainty and independent of the expected margin, predicting which firms err more on demand versus cost. Relative to unrestricted Rational Inattention (RI), the separate-channel structure also quantifies the value of informational integration across a firm's demand and cost functions.

Keywords: value of information, rational inattention, monopoly pricing, mutual information, directed information, demand uncertainty

Email address: `stefan@stefanbehringer.com` (Stefan Behringer)

JEL: D42, D81, D83, L12

1. Introduction

Information acquisition is central to monopoly pricing under demand uncertainty. The problem of how much a monopolist should pay for information about uncertain demand dates back at least to Marschak (1954), whose "monopolist" example derives the value of perfect demand information in a linear-quadratic setting and already observes that this value reduces to the variance of the demand state only because demand is linear in price. Continuous, variance-based analysis of information has since been a workhorse of economic theory, from team theory through the noisy rational-expectations tradition (Grossman and Stiglitz, 1980; Hellwig, 1980), in which signal precision is the natural parameter. The value of information itself is, in the Blackwell sense, monotone, more informative experiments are weakly more valuable, a fact Marschak (1954) already records as the non-negativity of his *value of inquiry* w, the gain in expected payoff from observing the state before acting ($w \geq 0$).

Against this background, Behringer (2021) offers a cautionary example: when demand is subject to both additive and multiplicative noise, the value of information can be *non-monotone in the noise variances.* Increasing the variance of multiplicative noise may raise it for certain parameter configurations, complicating comparative statics and welfare analysis. This is not a failure of learning, Blackwell monotonicity is untouched, but an artefact of using noise variance as the primitive when the signal structure is endogenous to the pricing decision: the same parameter that scales noise also reshapes the price-dependent signal, so movements in a variance need not correspond to monotone movements in information content. Such non-monotonicities do not arise with a single additive error or in static environments, which suggests the pathology lies in the parameterisation, not in the economics.

Subsequent work (Behringer and Belavkin, 2023, 2025) clarifies that this phenomenon is rooted in the endogeneity and circularity of information in pricing environments. Prices affect demand, observed demand affects beliefs, and beliefs feed back into future pricing decisions. In such settings, noise variances are poor proxies for information content: they conflate genuine learning about market fundamentals with mechanical amplification generated by the decision rule itself. As a result, comparative statics with respect to noise parameters need not reflect comparative statics with respect to information.

This observation connects naturally to the information-cost approach of Pomatto et al. (2023), who study information acquisition problems in which the primitive is the information content of an experiment rather than its physical noise structure. Under mild axiomsm, namely additivity across independent experiments and linearity in the probability of success, Pomatto et al. show that the admissible cost functions are exactly the log-likelihood-ratio (LLR) costs: a posterior-separable family that, unlike mutual information, is additive. From this perspective, non-monotonicities in the value of information cannot arise from learning per se; they can only reflect changes in the amount of information conveyed about the underlying state.

Our contribution is to bring this information-theoretic perspective to the monopoly pricing problem with endogenous feedback. First, we derive an explicit *mutual information formula* that maps Behringer's noise parameters $(\sigma_\varepsilon^2, \sigma_\gamma^2)$ into an information capacity $\kappa = I(\Theta, S)$. This mapping, which has not appeared in prior work, makes transparent how different combinations of additive and multiplicative noise translate into information content. We show that the apparent non-monotonicities in the value of information identified by Behringer (2021) arise precisely because changes in noise variances do not correspond to monotone changes in mutual information.

Second, we show that once information capacity κ is treated as the primitive parameter, rather than the underlying noise variances, the pathology disappears. The value of information is monotone and concave in information capacity, in line with the predictions of information-cost models à la Pomatto et al. This result holds even in environments with endogenous pricing, where feedback effects render standard signal-based approaches misleading.

Third, we let the firm face uncertainty in both demand and marginal cost, each learned through a separate channel, and characterise the optimal allocation of scarce attention between them. Because profit depends on the two unknowns only through their difference, the demand intercept net of marginal cost, the decision is one-dimensional while learning is two-dimensional, and the optimal split follows a *reverse water-filling rule* that tilts attention towards the more uncertain dimension, the reverse of ordinary water-filling, which favours the cleaner channel, specialising entirely when the asymmetry is large. Confronting the informationally siloed firm with the unrestricted RI benchmark of Sims (2003); Matějka and McKay (2015), in which the firm designs its information structure freely, identifies the resulting profit gap as the value of informational integration, an organisational-design object absent from single-signal models.

Our analysis builds on the Stratonovich/Shannon Value of Information (VoI) framework as laid out in Stratonovich (1975/2020), recently developed for economic applications in Behringer and Belavkin (2023, 2025) and, more generally, in Behringer (2026). Importantly, the latter framework does not rely on Gaussian signals, quadratic payoffs, or linear filtering, and establishes monotonicity and concavity of the VoI as fundamental properties of information itself. Our results can thus be viewed as a concrete implementation of this general theory in a monopoly pricing environment under uncertainty.

Finally, we introduce correlation, both between the additive and multiplicative errors, where it alters the mutual information bound, and between the demand and cost fundamentals, where it turns the two information channels into substitutes and reshapes the optimal allocation. In the first case the full filtering problem becomes analytically intractable, yet the information capacity remains tractable and yields sharp comparative statics. This reinforces the central message of the paper: when information is measured directly in terms of its content rather than its noise representation, the value of information is well behaved, monotone, and concave in both static and dynamic environments.

2. Behringer's (2021) Demand Model

2.1. Demand and Pricing

The demand faced by a Monopolist is given by

$$q = \theta - \varepsilon p + \gamma, \tag{1}$$

where $\theta \sim \mathcal{N}(\mu_0, \sigma_0^2)$ is market potential, drawn once, $\varepsilon \sim \mathcal{N}(1, \sigma_\varepsilon^2)$ is multiplicative noise, $\gamma \sim \mathcal{N}(0, \sigma_\gamma^2)$ is additive noise, drawn each period with mutual independence.

At unit cost c, monopoly profit is $\Pi = (p-c)q$. Given belief mean μ, the optimal price is

$$p^*(\mu) = \frac{\mu + c}{2}. \tag{2}$$

2.2. The Signal from Observed Demand

The firm sets price p based on current beliefs and observes realised demand q. Rearranging equation (1), the firm can construct the signal:

$$S = q + p = \theta - \varepsilon p + \gamma + p = \theta + p(1-\varepsilon) + \gamma. \tag{3}$$

Since $\varepsilon \sim \mathcal{N}(1, \sigma_\varepsilon^2)$, we have $1 - \varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2)$. Therefore:

$$S = \theta + \xi + \gamma, \tag{4}$$

where $\xi = p(1-\varepsilon) \sim \mathcal{N}(0, p^2\sigma_\varepsilon^2)$ is the price-dependent noise. Conditional on the state, the signal is therefore $S \mid \theta \sim \mathcal{N}\big(\theta,\ p^2\sigma_\varepsilon^2 + \sigma_\gamma^2\big)$. Note that this conditional (noise) variance $p^2\sigma_\varepsilon^2 + \sigma_\gamma^2$ is distinct from the prior variance σ_0^2 of θ itself; the two coincide only by coincidence.

2.3. Endogeneity and Heteroskedasticity

The key feature is that the signal variance depends on the chosen price:

$$\mathrm{Var}(S \mid \theta, p) = p^2\sigma_\varepsilon^2 + \sigma_\gamma^2. \tag{5}$$

When the firm uses optimal pricing $p^*(\mu_0) = \frac{\mu_0 + c}{2}$, this becomes:

$$\mathrm{Var}(S \mid \theta) = \frac{(\mu_0 + c)^2}{4}\sigma_\varepsilon^2 + \sigma_\gamma^2. \tag{6}$$

This creates two complications:

1. **Endogeneity**: Signal noise depends on the pricing decision.
2. **Heteroskedasticity**: Even conditional on θ, the effective noise varies with beliefs μ_0.

As a result, posterior variances and mutual information do not admit simple closed-form Gaussian expressions, and comparative statics with respect to $(\sigma_\varepsilon^2, \sigma_\gamma^2)$ become non-monotonic.

From this the non-monotonicity observed in Behringer (2021) may follow. In a multi-period setting, the value of information can *increase* with σ_ε^2 for certain parameter configurations. The mechanism is indirect: a higher σ_ε^2 raises the signal noise $p^2\sigma_\varepsilon^2$, but it also reshapes future pricing incentives and the signal structure, and these indirect effects can dominate, creating local non-monotonicities.

This pathology motivates our information-theoretic reformulation following the approach in Behringer and Belavkin (2023).

3. Mapping Noise Parameters into Information Capacity

3.1. The Information-Theoretic Approach

Behringer (2021) works with noise variances $(\sigma_\varepsilon^2, \sigma_\gamma^2)$ as parameters and derives the resulting information structure and posterior beliefs. His analysis shows that the signal variance is $\mathrm{Var}(S \mid \theta) = p^2\sigma_\varepsilon^2 + \sigma_\gamma^2$, where $p = \frac{\mu_0+c}{2}$ is the optimal price.

We build on this analysis by introducing a change of variables: rather than working with noise variances directly, we parameterise the problem by the *information capacity* κ, defined as the *Shannon mutual information* between the state Θ and the signal S:

$$\kappa \equiv I(\Theta, S). \tag{7}$$

This approach treats information content as primitive, making the value of information well-defined regardless of the specific noise structure. The key insight is that we derive the exact mutual information for Behringer's Gaussian quadratic model (not a bound), then use this to motivate treating κ as the primitive parameter.

3.2. Behringer's Model and the Mutual Information Formula

Behringer (2021) derives the posterior variance and Bayesian updating formulas but does not provide an explicit mutual information expression. We now derive this formula.

Proposition 1 (Mutual Information Formula for Behringer's Model)**.** *Given the demand model* (1) *with optimal pricing* $p^* = \frac{\mu_0+c}{2}$*, all variables jointly Gaussian and independent, the mutual information between state and signal is:*

$$I(\Theta, S) = \frac{1}{2}\log\left(1 + \frac{\sigma_0^2}{\frac{(\mu_0+c)^2}{4}\sigma_\varepsilon^2 + \sigma_\gamma^2}\right). \tag{8}$$

Proof. We establish this through the following calculations:

Step 1: Signal construction. From equation (3), the signal is:

$$S = \theta + p(1-\varepsilon) + \gamma, \tag{9}$$

where $p = \frac{\mu_0+c}{2}$ is the optimal price based on prior beliefs.

Step 2: Noise terms. Define $\xi = p(1-\varepsilon)$. Since $\varepsilon \sim \mathcal{N}(1, \sigma_\varepsilon^2)$:

$$E\{\xi\} = pE\{1-\varepsilon\} = p(1-1) = 0, \tag{10}$$

$$\mathrm{Var}(\xi) = p^2\mathrm{Var}(1-\varepsilon) = p^2\sigma_\varepsilon^2 = \frac{(\mu_0+c)^2}{4}\sigma_\varepsilon^2. \tag{11}$$

Since ξ and γ are independent (as ε and γ are independent):

$$S = \theta + \xi + \gamma, \quad \text{where } \xi \perp \gamma \perp \theta. \tag{12}$$

Step 3: Conditional variance. Given θ, the signal variance is:

$$\mathrm{Var}(S \mid \theta) = \mathrm{Var}(\xi) + \mathrm{Var}(\gamma) = \frac{(\mu_0+c)^2}{4}\sigma_\varepsilon^2 + \sigma_\gamma^2 \equiv \sigma_{\text{noise}}^2. \tag{13}$$

Step 4: Gaussianity. Since all variables are jointly Gaussian, (Θ, S) are jointly Gaussian.

Step 5: Posterior variance. The posterior distribution is:

$$\Theta \mid S \sim \mathcal{N}\left(\frac{\sigma_{\text{noise}}^2\mu_0 + \sigma_0^2 s}{\sigma_0^2 + \sigma_{\text{noise}}^2}, \frac{\sigma_0^2\sigma_{\text{noise}}^2}{\sigma_0^2 + \sigma_{\text{noise}}^2}\right). \tag{14}$$

Step 6: Computing MI. The mutual information for jointly Gaussian variables is, (e.g. Cover and Thomas (2006), p.263):

$$I(\Theta, S) = \frac{1}{2}\log\frac{\sigma_0^2}{\sigma_1^2} = \frac{1}{2}\log\left(1 + \frac{\sigma_0^2}{\sigma_{\text{noise}}^2}\right) = \frac{1}{2}\log\left(1 + \frac{\sigma_0^2}{\frac{(\mu_0+c)^2}{4}\sigma_\varepsilon^2 + \sigma_\gamma^2}\right). \tag{15}$$

$\square$

3.3. Extension: Correlated Errors

We now consider the case where the multiplicative and additive errors are correlated.

Let $\varepsilon \sim \mathcal{N}(1, \sigma_\varepsilon^2)$ and $\gamma \sim \mathcal{N}(0, \sigma_\gamma^2)$ with correlation coefficient $\rho = \text{Corr}(\varepsilon, \gamma)$. The covariance is:

$$\text{Cov}(\varepsilon, \gamma) = \rho\sigma_\varepsilon\sigma_\gamma. \tag{16}$$

Proposition 2 (MI with Correlated Errors)**.** *With correlated errors, the mutual information is:*

$$I(\Theta, S) = \frac{1}{2}\log\left(1 + \frac{\sigma_0^2}{\frac{(\mu_0+c)^2}{4}\sigma_\varepsilon^2 + \sigma_\gamma^2 - (\mu_0 + c)\rho\sigma_\varepsilon\sigma_\gamma}\right). \tag{17}$$

Proof. The signal is still $S = \theta + p(1 - \varepsilon) + \gamma$ where $p = \frac{\mu_0+c}{2}$.

Define $\xi = p(1 - \varepsilon)$. We have:

$$E\{\xi\} = pE\{1 - \varepsilon\} = 0, \tag{18}$$

$$\text{Var}(\xi) = p^2\text{Var}(\varepsilon) = p^2\sigma_\varepsilon^2. \tag{19}$$

The covariance between ξ and γ is:

$$\text{Cov}(\xi, \gamma) = \text{Cov}(p(1-\varepsilon), \gamma) = p\cdot\text{Cov}(1-\varepsilon, \gamma) = -p\cdot\text{Cov}(\varepsilon, \gamma) = -p\rho\sigma_\varepsilon\sigma_\gamma. \tag{20}$$

The total noise is $N = \xi + \gamma$, which has variance:

$$\mathrm{Var}(N) = \mathrm{Var}(\xi) + \mathrm{Var}(\gamma) + 2\mathrm{Cov}(\xi, \gamma) \tag{21}$$

$$= p^2\sigma_\varepsilon^2 + \sigma_\gamma^2 + 2(-p\rho\sigma_\varepsilon\sigma_\gamma) \tag{22}$$

$$= p^2\sigma_\varepsilon^2 + \sigma_\gamma^2 - 2p\rho\sigma_\varepsilon\sigma_\gamma. \tag{23}$$

Substituting $p = \frac{\mu_0+c}{2}$:

$$\mathrm{Var}(N) = \frac{(\mu_0 + c)^2}{4}\sigma_\varepsilon^2 + \sigma_\gamma^2 - (\mu_0 + c)\rho\sigma_\varepsilon\sigma_\gamma. \tag{24}$$

Since the signal is still Gaussian (linear combination of joint Gaussians), the mutual information is:

$$I(\Theta, S) = \frac{1}{2}\log\left(1 + \frac{\sigma_0^2}{\mathrm{Var}(N)}\right) = \frac{1}{2}\log\left(1 + \frac{\sigma_0^2}{\frac{(\mu_0+c)^2}{4}\sigma_\varepsilon^2 + \sigma_\gamma^2 - (\mu_0 + c)\rho\sigma_\varepsilon\sigma_\gamma}\right). \tag{25}$$

□

Corollary 1 (Effect of Correlation)**.** *The effect of correlation on mutual information is:*

1. *If $\rho > 0$ (positive correlation): Var(N) decreases, so $I(\Theta; S)$* ***increases***
2. *If $\rho < 0$ (negative correlation): Var(N) increases, so $I(\Theta; S)$* ***decreases***
3. *If $\rho = 0$ (independent): reduces to equation* (8)

Intuitively, positive correlation between ε and γ means that when the multiplicative shock is high (demand more sensitive to price), the additive shock tends to be positive (higher demand). These effects partially cancel in the signal $S = \theta + p(1 - \varepsilon) + \gamma$, reducing total noise variance and increasing information content. This formalises, in the monopoly setting, the conjecture of Marschak (1954, §5.7) that statistical dependence between observables alters the value of learning about them.

While the mutual information formula extends cleanly to correlated errors, the full Bayesian filtering process becomes significantly more complex. The posterior distribution $\Theta \mid S$ remains Gaussian, but updating the posterior covariance structure across multiple periods with correlated noise requires matrix

Kalman-Stratonovich filtering. For tractability in dynamic settings, we therefore focus on the independent case $\rho = 0$ in the remainder of the paper.

In either case, equation (8) provides a rigorous mapping from the noise parameters $(\sigma_\varepsilon^2, \sigma_\gamma^2)$ to the information capacity κ, exact whenever all variables are jointly Gaussian.

3.4. Induced Information Capacity

Definition 1 (Induced Information Capacity)**.** *Given noise parameters* $(\sigma_\varepsilon^2, \sigma_\gamma^2)$ *and prior beliefs* (μ_0, σ_0^2)*, the* induced information capacity *is:*

$$\kappa(\sigma_\varepsilon^2, \sigma_\gamma^2) \equiv I(\Theta, S), \tag{26}$$

where S is the signal constructed from observed demand under optimal pricing.

Proposition 3 (Monotonicity of Induced Capacity)**.** *The induced information capacity $\kappa(\sigma_\varepsilon^2, \sigma_\gamma^2)$ is strictly decreasing in both arguments.*

Proof. From equation (8), κ is a decreasing function of the total noise variance $\frac{(\mu_0+c)^2}{4}\sigma_\varepsilon^2 + \sigma_\gamma^2$. Increasing either σ_ε^2 or σ_γ^2 increases this variance, reducing mutual information by the *data processing inequality.* (e.g. Cover and Thomas (2006), p.34) □

While κ is monotone in noise parameters, the value of information in Behringer's model can be non-monotone because changing $(\sigma_\varepsilon^2, \sigma_\gamma^2)$ affects both κ (via equation (8)) and the pricing strategy in complex ways. The VoI approach of Behringer and Belavkin (2023, 2025) resolves this by treating κ as the primitive parameter.[1]

[1]The formulation of VoI is in terms of signal structures $p(s|\theta)$ subject to $I(\Theta, S) \le \kappa$ is equivalent to optimising over action policies $q(s)$. By the Markov chain $\Theta \to S \to Q$ and the *data processing inequality* (see above), we have $I(\Theta, Q) \le I(\Theta, S)$, with equality when $Q = q^*(S)$ is the optimal deterministic action. Thus, optimising the signal structure is equivalent to choosing what information to acquire, with the optimal action following automatically via Bayesian updating.

4. Canonical Model with Exogenous Information Capacity

4.1. Setup

We now formulate a *canonical* model where information capacity κ is specified exogenously, and the noise structure is derived endogenously.

Market structure: Given the market potential prior: $\theta \sim \mathcal{N}(\mu_0, \sigma_0^2)$, demand is $q = \theta - p$ (deterministic conditional on θ) and constant marginal cost c. Thus profit is: $\pi = (p - c)q$.

Information acquisition: The firm can observe a signal s about θ subject to:

$$I(\Theta, S) \leq \kappa, \tag{27}$$

where $\kappa \geq 0$ is the exogenous information capacity.

4.2. Optimal Signal Structure

Theorem 1 (Optimal Signal for Gaussian Prior)**.** *For Gaussian prior $\theta \sim \mathcal{N}(\mu_0, \sigma_0^2)$, the optimal signal structure achieving $I(\Theta, S) = \kappa$ is additive Gaussian:*

$$s = \theta + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2(\kappa)), \tag{28}$$

where

$$\sigma_\varepsilon^2(\kappa) = \frac{\sigma_0^2}{e^{2\kappa} - 1}. \tag{29}$$

Proof. For jointly Gaussian variables, mutual information is:

$$I(\Theta, S) = \frac{1}{2} \log \frac{\sigma_0^2}{\sigma_1^2}, \tag{30}$$

where $\sigma_1^2 = \mathrm{Var}(\Theta \mid S)$ is the posterior variance. With additive Gaussian noise $s = \theta + \varepsilon$:

$$\sigma_1^2 = \left(\frac{1}{\sigma_0^2} + \frac{1}{\sigma_\varepsilon^2} \right)^{-1} = \frac{\sigma_0^2 \sigma_\varepsilon^2}{\sigma_0^2 + \sigma_\varepsilon^2}. \tag{31}$$

Setting $I(\Theta, S) = \kappa$:

$$e^{2\kappa} = \frac{\sigma_0^2}{\sigma_1^2} = \frac{\sigma_0^2 + \sigma_\varepsilon^2}{\sigma_\varepsilon^2}, \tag{32}$$

which yields equation (29). □

Corollary 2 (Posterior Variance)**.** *Under information constraint κ, the posterior variance is:*

$$\sigma_1^2(\kappa) = \sigma_0^2 e^{-2\kappa}. \tag{33}$$

4.3. Learning Process

We now describe the Bayesian learning process under the information capacity constraint, paralleling the development in Behringer (2021). Throughout, σ_0^2 denotes the initial prior variance and σ_t^2 the posterior variance after period t, so the symbol σ_1^2 always refers to a posterior, consistent with the corollary above and with Section 5 below.

4.3.1. Single-Period Learning

Prior beliefs: The firm holds the prior $\theta \sim \mathcal{N}(\mu_0, \sigma_0^2)$.

Signal: It observes $s_1 = \theta + \varepsilon_1$ with $\varepsilon_1 \sim \mathcal{N}(0, \sigma_\varepsilon^2(\kappa_1))$ and $I(\Theta, S_1) = \kappa_1$, where the additive noise achieving capacity κ_1 against a Gaussian prior of variance σ_0^2 is $\sigma_\varepsilon^2(\kappa_1) = \sigma_0^2/(e^{2\kappa_1} - 1)$ from (29). (For a later signal the same expression applies with the current prior variance in place of σ_0^2.)

Posterior beliefs: Bayesian updating for Gaussian distributions gives the posterior mean

$$(1 - \lambda_1)\mu_0 + \lambda_1 s_1, \qquad \lambda_1 = \frac{\sigma_0^2}{\sigma_0^2 + \sigma_\varepsilon^2(\kappa_1)} = 1 - e^{-2\kappa_1}, \tag{34}$$

and the posterior variance $\sigma_0^2 e^{-2\kappa_1}$, in agreement with the corollary.

4.3.2. Multi-Period Learning with Demand Observations

Consider a two-period setting as in Behringer (2021).

Period 1:

1. Prior: $\theta \sim \mathcal{N}(\mu_0, \sigma_0^2)$.
2. Observe signal: $s_1 = \theta + \varepsilon_1$ with $I(\Theta, S_1) = \kappa_1$.
3. Set price: $p_1 = \frac{\mu_0 + c}{2}$ (based on the prior).
4. Observe demand: $q_1 = \theta - p_1 + \eta_1$ where $\eta_1 \sim \mathcal{N}(0, \sigma_\eta^2)$ is transitory noise.
5. Construct second signal: $D_1 = q_1 + p_1 = \theta + \eta_1$.

Updating after Period 1:

The firm now has two independent signals about θ:

$$s_1 = \theta + \varepsilon_1, \quad \varepsilon_1 \sim \mathcal{N}(0, \sigma_\varepsilon^2(\kappa_1)) \tag{35}$$

$$D_1 = \theta + \eta_1, \quad \eta_1 \sim \mathcal{N}(0, \sigma_\eta^2) \tag{36}$$

By Kalman-Stratonovich filtering (or standard Bayesian updating with multiple signals),

$$\frac{1}{\sigma_1^2} = \frac{1}{\sigma_0^2} + \frac{1}{\sigma_\varepsilon^2(\kappa_1)} + \frac{1}{\sigma_\eta^2}, \tag{37}$$

where σ_1^2 denotes the posterior variance after period 1. Using $\frac{1}{\sigma_\varepsilon^2(\kappa_1)} = \frac{e^{2\kappa_1}-1}{\sigma_0^2}$,

$$\frac{1}{\sigma_1^2} = \frac{1}{\sigma_0^2} + \frac{e^{2\kappa_1}-1}{\sigma_0^2} + \frac{1}{\sigma_\eta^2} = \frac{e^{2\kappa_1}}{\sigma_0^2} + \frac{1}{\sigma_\eta^2}, \tag{38}$$

so that

$$\sigma_1^2 = \left(\frac{e^{2\kappa_1}}{\sigma_0^2} + \frac{1}{\sigma_\eta^2} \right)^{-1}. \tag{39}$$

The posterior mean is

$$\mu_1 = \sigma_1^2 \left(\frac{\mu_0}{\sigma_0^2} + \frac{s_1}{\sigma_\varepsilon^2(\kappa_1)} + \frac{D_1}{\sigma_\eta^2} \right). \tag{40}$$

Period 2:

6. Prior: $\theta \sim \mathcal{N}(\mu_1, \sigma_1^2)$ (beliefs from period 1).
7. Observe signal: $s_2 = \theta + \varepsilon_2$ with $I(\Theta; S_2 \mid \mathcal{I}_1) = \kappa_2$, where $\mathcal{I}_1 = (s_1, D_1)$ is the period-1 information set.
8. Set price: $p_2 = \frac{\mu_1 + c}{2}$ (based on updated beliefs).
9. Realise profit π_2.

Final posterior:

$$\sigma_2^2 = \sigma_1^2 e^{-2\kappa_2} = \left(\frac{e^{2\kappa_1}}{\sigma_0^2} + \frac{1}{\sigma_\eta^2} \right)^{-1} e^{-2\kappa_2}. \tag{41}$$

Proposition 4 (Information Accumulation)**.** *The total information accumulated over two periods is characterised by the reduction in posterior variance:*

$$\frac{\sigma_0^2}{\sigma_2^2} = \left(\frac{e^{2\kappa_1}}{\sigma_0^2} + \frac{1}{\sigma_\eta^2} \right) \sigma_0^2 \cdot e^{2\kappa_2} = e^{2\kappa_2} \left(e^{2\kappa_1} + \frac{\sigma_0^2}{\sigma_\eta^2} \right). \tag{42}$$

Equivalently, in terms of posterior precision,

$$\frac{1}{\sigma_2^2} = \frac{e^{2\kappa_2}}{\sigma_0^2}\left(e^{2\kappa_1} + \frac{\sigma_0^2}{\sigma_\eta^2}\right) = \frac{e^{2(\kappa_1+\kappa_2)}}{\sigma_0^2} + \frac{e^{2\kappa_2}}{\sigma_\eta^2}. \tag{43}$$

When demand observations are uninformative ($\sigma_\eta^2 \to \infty$), the transitory channel vanishes and the two signal capacities accumulate additively: posterior precision grows as $\frac{1}{\sigma_2^2} \to \frac{e^{2(\kappa_1+\kappa_2)}}{\sigma_0^2}$, i.e. the capacities add to $\kappa_1 + \kappa_2$. In the opposite limit of perfect demand observations ($\sigma_\eta^2 \to 0$), the term $e^{2\kappa_2}/\sigma_\eta^2 \to \infty$, so $\sigma_2^2 \to 0$: the demand observation fully reveals θ and the signal capacities become irrelevant.

The clean additive form is thus the no-demand-information benchmark: it isolates the contribution of the two capacity-constrained signals. Informative demand observations ($\sigma_\eta^2 < \infty$) only accelerate learning, adding the extra precision $e^{2\kappa_2}/\sigma_\eta^2$ on top, so the posterior precision is everywhere weakly above the additive benchmark and strictly above it whenever σ_η^2 is finite.

Behringer (2021) derives updating formulas with explicit noise parameters $(\sigma_\varepsilon^2, \sigma_\gamma^2)$. Our formulation achieves the same learning dynamics but parameterised by information capacities κ_1, κ_2. The key difference is that κ directly measures information content, making comparative statics more transparent.

4.4. Optimal Quantity Choice

Given signal s, the posterior mean is:

$$\mu(s) = E\{\theta \mid s\} = \frac{\sigma_\varepsilon^2}{\sigma_0^2 + \sigma_\varepsilon^2}\mu_0 + \frac{\sigma_0^2}{\sigma_0^2 + \sigma_\varepsilon^2}s. \tag{44}$$

Define the learning weight:

$$\lambda(\kappa) = \frac{\sigma_0^2}{\sigma_0^2 + \sigma_\varepsilon^2} = 1 - e^{-2\kappa}. \tag{45}$$

Then:

$$\mu(s) = (1-\lambda)\mu_0 + \lambda s. \tag{46}$$

The optimal quantity maximises expected profit $E\{(\theta - q - c)q \mid s\}$:

$$q^*(s) = \frac{\mu(s) - c}{2}. \tag{47}$$

4.5. Value Function

Expected profit conditional on signal s:

$$E\{\pi \mid s\} = E\{(\theta - q^* - c)q^* \mid s\} = \frac{(\mu(s) - c)^2}{4}. \tag{48}$$

Ex-ante expected profit (before observing signal):

$$V(\kappa) = E_s\left\{\frac{(\mu(s) - c)^2}{4}\right\} = \frac{1}{4}E_s\{(\mu(s) - c)^2\}. \tag{49}$$

Since $s \sim \mathcal{N}(\mu_0, \sigma_0^2 + \sigma_\varepsilon^2)$ and $\mu(s) = (1 - \lambda)\mu_0 + \lambda s$:

$$E\{(\mu(s) - c)^2\} = \mathrm{Var}[\mu(s)] + (E\{\mu(s)\} - c)^2 \tag{50}$$

$$= \lambda^2(\sigma_0^2 + \sigma_\varepsilon^2) + (\mu_0 - c)^2. \tag{51}$$

Using $\sigma_0^2 + \sigma_\varepsilon^2 = \sigma_0^2 \frac{e^{2\kappa}}{e^{2\kappa}-1}$ and $\lambda = 1 - e^{-2\kappa}$:

$$\mathrm{Var}[\mu(s)] = (1 - e^{-2\kappa})^2 \cdot \sigma_0^2 \frac{e^{2\kappa}}{e^{2\kappa} - 1} \tag{52}$$

$$= \frac{(e^{2\kappa} - 1)^2}{e^{4\kappa}} \cdot \sigma_0^2 \frac{e^{2\kappa}}{e^{2\kappa} - 1} \tag{53}$$

$$= \sigma_0^2 \frac{e^{2\kappa} - 1}{e^{2\kappa}} = \sigma_0^2(1 - e^{-2\kappa}). \tag{54}$$

Therefore:

$$V(\kappa) = \frac{(\mu_0 - c)^2 + \sigma_0^2(1 - e^{-2\kappa})}{4}. \tag{55}$$

4.6. Value of Information

With no information ($\kappa = 0$), expected profit is:

$$V(0) = \frac{(\mu_0 - c)^2}{4}. \tag{56}$$

The Value of Information (VoI) is:

$$\mathrm{VoI}(\kappa) = V(\kappa) - V(0) = \frac{\sigma_0^2}{4}(1 - e^{-2\kappa}). \tag{57}$$

Theorem 2 (Properties of VoI)**.** *The Value of Information function satisfies:*

1. ***Normalisation:*** $VoI(0) = 0$
2. ***Monotonicity:*** $\frac{dVoI}{d\kappa} = \frac{\sigma_0^2}{2}e^{-2\kappa} > 0$ *for all* $\kappa \geq 0$

3. ***Concavity:*** $\frac{d^2\,VoI}{d\kappa^2} = -\sigma_0^2 e^{-2\kappa} < 0$ *for all* $\kappa \geq 0$
4. ***Bounded:*** $\lim_{\kappa\to\infty} VoI(\kappa) = \frac{\sigma_0^2}{4}$

Proof. Direct differentiation of the VoI formula. □

The concavity in part (3) is not special to the Gaussian-quadratic case. Whitmeyer (2025) shows quite generally that when the amount of information is quantified as a convex divergence of posteriors and acquired efficiently, the resulting value of information is globally concave in that amount, with strictly positive marginal value at zero. Our closed form is a parametric instance of this principle, with mutual information κ as the divergence-based measure of amount.

The full-information limit $\lim_{\kappa\to\infty} \mathrm{VoI}(\kappa) = \sigma_0^2/4$ recovers the classical result of Marschak (1954), who showed that the value of perfect demand information to a monopolist equals the variance of the demand state (here scaled by the $q = \theta - p$ normalisation).

Remark 1 (Mutual information as a sufficient statistic). Within the scalar Gaussian additive family $s = \theta + \varepsilon$ used throughout, $\kappa = \frac{1}{2}\log(\sigma_0^2/\sigma_1^2)$ is strictly monotone in the posterior variance $\sigma_1^2 = \sigma_0^2 e^{-2\kappa}$. Hence two signals in this family are Blackwell-equivalent if and only if they share the same κ, which is what justifies treating κ as the sole informational primitive. Across arbitrary experiments this equivalence fails: equal mutual information is necessary but not sufficient for Blackwell equivalence, since the scalar κ cannot order experiments that the (partial) Blackwell order leaves incomparable.

4.7. Special Role of Gaussian-Quadratic Setting

The particularly simple closed-form VoI in equation (57) is a special case of the Gibbs (Stratonovich) structure underlying the general framework of Behringer and Belavkin (2023, 2025); Behringer (2026). There, the information-optimal signal channel takes the exponential-family form

$$p(s \mid \theta) = \frac{p_0(s)}{Z(\theta,\beta)} \exp(\beta\, U(a^*(s),\theta)), \qquad Z(\theta,\beta) = \int p_0(s) \exp(\beta\, U(a^*(s),\theta))\, ds, \tag{58}$$

where $p_0(s)$ is a baseline reference measure over signals, $Z(\theta, \beta)$ is the (in general state-dependent) partition function, $\beta = 1/\lambda$ is the inverse information *temperature*, and the free energy $F(\theta) = -\lambda \log Z(\theta, \beta)$ captures nonlinear effects. The Gaussian-quadratic (monopoly profit) setting collapses this machinery to closed form through three special features, all fully relaxed in Behringer and Belavkin (2025); Behringer (2026):

1. **Exponential family preservation:** Completing the square, monopoly profit separates into a state-only term and a quadratic loss around the full-information quantity $q^\dagger(\theta) = \frac{\theta - c}{2}$:

$$\pi(q, \theta) = \frac{(\theta - c)^2}{4} - \left(q - q^\dagger(\theta)\right)^2. \tag{59}$$

   Because the action-dependent part is quadratic in the gap $q - q^\dagger(\theta)$, the exponential of profit factors as

$$\exp(\beta\, \pi(q, \theta)) = \underbrace{e^{\beta(\theta - c)^2/4}}_{\text{state-only}} \exp\Big(-\beta \left(q - q^\dagger(\theta)\right)^2\Big), \tag{60}$$

   which is Gaussian in q. Here β is the inverse information temperature (the multiplier of the underlying Lagrangian), and the Gibbs channel $p(s \mid \theta) \propto \exp(\beta\, \pi(q^*(s), \theta))$ is therefore Gaussian as well.

2. **State-independent partition function (for the loss):** The action-dependent part of profit depends only on the gap $q - q^\dagger(\theta)$, i.e. it is *translation invariant*. With a location-family signal $s = \theta + \varepsilon$, the change of variables $\varepsilon = s - \theta$ removes the state from the integral, so the loss partition function

$$Z(\beta) = \int_{\mathcal{S}} \exp\Big(-\beta \left(q^*(s) - q^\dagger(\theta)\right)^2\Big)\, ds \tag{61}$$

   is independent of θ. State-independence is thus a *translation-invariance* property: any shift-invariant loss would deliver it. The explicit value $Z(\beta) = \sqrt{\pi/\beta}$ then uses the quadratic form (a Gaussian integral). The state-only factor $e^{\beta(\theta - c)^2/4}$ is separable and does not affect the optimal action.

3. **Certainty equivalence:** With a quadratic objective, the optimal quantity depends only on the posterior mean, $q^*(s) = \frac{\mu(s)-c}{2}$. Completing the square, monopoly profit decomposes as $\pi = \frac{(\theta-c)^2}{4} - \left(q - \frac{\theta-c}{2}\right)^2$, so the gross term and the posterior loss exactly cancel:

$$E\{\pi \mid s\} = \frac{(\mu(s)-c)^2 + \sigma_1^2}{4} - \frac{\sigma_1^2}{4} = \frac{(\mu(s)-c)^2}{4}, \tag{62}$$

leaving conditional expected profit a function of the posterior mean alone, consistent with the value function of Section 4.5. The posterior variance σ_1^2 thus governs the *ex-ante* dispersion of $\mu(s)$ (hence the value of information) without entering the conditional profit directly.

The full signal distribution beyond these moments is unnecessary.

That variance reduction is a sufficient statistic for the VoI here is not a generic feature but a knife-edge property of the linear-quadratic-Gaussian structure. This was already anticipated by Marschak (1954), who noted that the value of demand information collapses to the variance σ^2 precisely *because* demand is linear in price, and that absent linearity some other parameter of the distribution would be the relevant measure of variability; the point is sharpened in Behringer (2026), where variance-sufficiency of the VoI is shown to require quadratic payoffs, a Gaussian prior, and conjugacy-preserving signals simultaneously.

In the more general case, the VoI may no longer reveal a closed form. The partition function $Z(\theta, \beta)$ depends on the state θ. The Markov chain argument (see footnote 1) still holds, ensuring VoI depends only on κ, not the specific action rule. The partition function $Z(\theta, \beta)$ must typically be evaluated numerically, and the VoI obtained via parametric plots relating $U(\beta)$ and $I(\beta)$ as in Behringer and Belavkin (2023).

The Gaussian-quadratic case studied in this paper is thus a special "degenerate" case where the full thermodynamic structure is present but hidden, yielding tractable closed-form solutions.

5. Directed Information and Dynamic Learning

5.1. Motivation: Feedback and Circularity

In multi-period settings, prices affect demand, which affects future beliefs and pricing. Standard mutual information $I(\Theta, D^T)$ cannot properly account for this feedback because it treats the entire demand sequence symmetrically.

Directed information resolves this by measuring information flow in causal order, conditioning on past observations and is attributed to Massey (1990).

5.2. Directed Information

Definition 2 (Directed Information). *The directed information from Θ to a process $D^T = (D_1, \ldots, D_T)$ is:*

$$I(\Theta \to D^T) = \sum_{t=1}^{T} I(\Theta, D_t \mid D^{t-1}), \tag{63}$$

where $D^{t-1} = (D_1, \ldots, D_{t-1})$ denotes the history.

This measures the cumulative information about Θ revealed across time, properly accounting for conditioning on past observations.

For a fixed fundamental Θ, the chain rule for mutual information gives $\sum_{t=1}^{T} I(\Theta; D_t \mid D^{t-1}) = I(\Theta; D^T)$, so directed and ordinary mutual information coincide here: when the source is a single, time-invariant state, nothing is lost by the symmetric measure. The directed formulation departs from mutual information precisely when the source itself evolves in response to past observations or actions, that is, under genuine feedback. The causal, period-by-period decomposition is what makes it the right accounting device once today's price shapes tomorrow's state or signal, which is why we adopt it as the natural language for the dynamic problem even though, in the two-period Gaussian model below, it reduces to the mutual-information bookkeeping already used.

5.3. Two-Period Model

Period 1:

1. Prior: $\theta \sim \mathcal{N}(\mu_0, \sigma_0^2)$
2. Observe signal $s_1 = \theta + \varepsilon_1$ with $I(\Theta, S_1) = \kappa_1$
3. Choose $q_1(s_1) = \frac{\mu(s_1)-c}{2}$
4. Observe noisy demand: $D_1 = \theta + \eta_1$ where $\eta_1 \sim \mathcal{N}(0, \sigma_\eta^2)$
5. Update beliefs: $\theta \mid s_1, D_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$

Period 2:

6. Observe signal $s_2 = \theta + \varepsilon_2$ with $I(\Theta, S_2 \mid \mathcal{I}_1) = \kappa_2$
7. Choose $q_2(s_2, \mathcal{I}_1)$
8. Realise profit π_2

Here $\mathcal{I}_1 = (s_1, D_1)$ is the period-1 information set.

5.4. Information Accumulation via Kalman-Stratonovich Filtering

After period 1, the firm has observed two independent signals:

$$s_1 = \theta + \varepsilon_1, \quad \varepsilon_1 \sim \mathcal{N}(0, \sigma_{\varepsilon,1}^2(\kappa_1)) \tag{64}$$

$$D_1 = \theta + \eta_1, \quad \eta_1 \sim \mathcal{N}(0, \sigma_\eta^2). \tag{65}$$

By standard Kalman-Stratonovich filtering, the posterior precision is:

$$\frac{1}{\sigma_1^2} = \frac{1}{\sigma_0^2} + \frac{1}{\sigma_{\varepsilon,1}^2} + \frac{1}{\sigma_\eta^2}. \tag{66}$$

Using $\frac{1}{\sigma_{\varepsilon,1}^2} = \frac{e^{2\kappa_1}-1}{\sigma_0^2}$:

$$\frac{1}{\sigma_1^2} = \frac{1}{\sigma_0^2} + \frac{e^{2\kappa_1}-1}{\sigma_0^2} + \frac{1}{\sigma_\eta^2} = \frac{e^{2\kappa_1}}{\sigma_0^2} + \frac{1}{\sigma_\eta^2}. \tag{67}$$

Therefore:

$$\sigma_1^2 = \left(\frac{e^{2\kappa_1}}{\sigma_0^2} + \frac{1}{\sigma_\eta^2}\right)^{-1}. \tag{68}$$

In period 2, with new signal capacity κ_2:

$$\sigma_2^2 = \sigma_1^2 e^{-2\kappa_2}. \tag{69}$$

Period 2 value (conditional on the posterior variance σ_1^2carried in):

$$V_2(\sigma_1^2, \kappa_2) = \frac{E\{(\mu_1 - c)^2\} + \sigma_1^2(1 - e^{-2\kappa_2})}{4}. \tag{70}$$

Period 1 flow value (the static value at capacity κ_1):

$$V_1(\kappa_1) = \frac{(\mu_0 - c)^2 + \sigma_0^2(1 - e^{-2\kappa_1})}{4}. \tag{71}$$

Taking expectations over the period-1 signals and using the law of iterated expectations,

$$E\{(\mu_1 - c)^2\} = (\mu_0 - c)^2 + \mathrm{Var}(\mu_1) = (\mu_0 - c)^2 + \sigma_0^2 - \sigma_1^2, \tag{72}$$

where $\sigma_1^2 = \sigma_1^2(\kappa_1) = \left(e^{2\kappa_1}/\sigma_0^2 + 1/\sigma_\eta^2\right)^{-1}$ is the posterior variance entering period 2.

The total discounted value is $V_{\text{total}} = V_1 + \delta\, E\{V_2\}$, with $\delta \in (0, 1)$ the discount factor:

$$\begin{aligned} V_{\text{total}}(\kappa_1, \kappa_2) = {} & \frac{(\mu_0 - c)^2}{4} + \frac{\sigma_0^2}{4}(1 - e^{-2\kappa_1}) \\ & + \delta \left[\frac{(\mu_0 - c)^2 + \sigma_0^2 - \sigma_1^2}{4} + \frac{\sigma_1^2}{4}(1 - e^{-2\kappa_2}) \right]. \end{aligned} \tag{73}$$

Summing the discounted single-period signal values isolates the direct contribution of the two capacity-constrained signals,

$$\mathrm{VoI}_{\text{sig}}(\kappa_1, \kappa_2) = \frac{\sigma_0^2}{4} \left[(1 - e^{-2\kappa_1}) + \delta \frac{\sigma_1^2}{\sigma_0^2}(1 - e^{-2\kappa_2}) \right]. \tag{74}$$

This is the dynamic counterpart of the static $\mathrm{VoI}(\kappa) = V(\kappa) - V(0)$, but it understates the full information premium $V_{\text{total}}(\kappa_1, \kappa_2) - V_{\text{total}}(0, 0)$ by $\frac{\delta}{4}\left(\bar{\sigma}_1^2 - \sigma_1^2(\kappa_1)\right) \geq 0$, where $\bar{\sigma}_1^2 = \sigma_1^2|_{\kappa_1 = 0} = \left(1/\sigma_0^2 + 1/\sigma_\eta^2\right)^{-1}$: early capacity κ_1 additionally raises period-2 base profit by sharpening the prior carried into period 2, an intertemporal benefit beyond the two signals' direct values.

Proposition 5 (Dynamic Monotonicity, Concavity, and Intertemporal Substitutability)**.** *The total discounted value $V_{total}(\kappa_1, \kappa_2)$ is strictly increasing in* *both κ_1 and κ_2 and strictly concave in κ_2. The two capacities are intertemporal* substitutes*:*

$$\frac{\partial^2 V_{total}}{\partial \kappa_1 \partial \kappa_2} = \frac{\delta}{2}\, e^{-2\kappa_2}\, \frac{d\sigma_1^2(\kappa_1)}{d\kappa_1} < 0, \qquad \sigma_1^2(\kappa_1) = \left(\frac{e^{2\kappa_1}}{\sigma_0^2} + \frac{1}{\sigma_\eta^2} \right)^{-1}, \tag{75}$$

since $\sigma_1^2(\kappa_1)$ is strictly decreasing in κ_1.

Proof. Increasingness in each argument and concavity in κ_2 follow by direct differentiation. The marginal value of period-2 capacity, $\partial V_{\text{total}}/\partial \kappa_2 = \frac{\delta}{2}\, \sigma_1^2(\kappa_1)\, e^{-2\kappa_2}$, is proportional to the residual uncertainty $\sigma_1^2(\kappa_1)$ carried into period 2. Because higher early capacity κ_1 lowers $\sigma_1^2(\kappa_1)$, it lowers the marginal value of κ_2: early and late information are substitutes, not complements. □

The value of information is monotone and concave in capacity, exactly the benefit a rationally inattentive firm weighs against the cost of acquiring it, and the analysis now turns from the value of information to its optimal acquisition in the tradition of Sims (2003) and Matějka and McKay (2015). The recasting in terms of mutual and directed information is not itself the economic content but the device that makes it tractable: a noise-variance formulation cannot even pose the acquisition problem cleanly, since the parameter that scales noise simultaneously reshapes the price-dependent signal, so how much to learn has no parameter-free answer. Capacity supplies one, and pricing under demand uncertainty is precisely the environment Denti (2022) singles out as a leading case for the posterior-separable information costs on which this approach rests.

6. Optimal Information Acquisition

6.1. Information Acquisition Costs

Information has a cost, and a natural discipline on it is constant marginal cost in production. Pomatto et al. (2023) show that the cost functions additive across independent experiments and linear in the probability of success are exactly the log-likelihood-ratio (LLR) costs $C(\mu) = \sum_{i,j} \beta_{ij}\, D_{\text{KL}}(\mu_i \,\|\, \mu_j)$, a posterior-separable family that is *additive*, in contrast to the subadditive mutual-information cost of the later RI literature. We summarise information by the scalar capacity $\kappa = I(\Theta, S)$, a monotone sufficient statistic for the posterior in the Gaussian family used here, and take the acquisition cost to be a convex increasing $C(\kappa)$ with $C(0) = 0$, $C'(\kappa) > 0$, and $C''(\kappa) > 0$, an increasing transformation of mutual information in the sense of Denti (2022).[2] This class does not abandon the LLR cost but contains it: specialised to a normal measurement the LLR cost is proportional to posterior precision (Pomatto et al., 2023, Prop. 2), so with $1/\sigma_\varepsilon^2(\kappa) = (e^{2\kappa} - 1)/\sigma_0^2$ it reads $C_{\text{LLR}}(\kappa) = a\,(e^{2\kappa} - 1)$ for some $a > 0$, convex, increasing, and zero at $\kappa = 0$. The linear specification $C(\kappa) = \lambda\kappa$ is the cost of the later RI literature.

The firm's net payoff from acquiring information is its VoI less the acquisition cost,

$$\Pi^{\text{net}}(\kappa) = \text{VoI}(\kappa) - C(\kappa) = \frac{\sigma_0^2}{4}(1 - e^{-2\kappa}) - C(\kappa), \tag{76}$$

and the firm chooses κ to maximise it. The first-order condition equates the marginal value of information to its marginal cost,

$$\left.\frac{d\text{VoI}}{d\kappa}\right|_{\kappa^*} = C'(\kappa^*) \quad \Longleftrightarrow \quad \frac{\sigma_0^2}{2} e^{-2\kappa^*} = C'(\kappa^*). \tag{77}$$

Linear RI cost. With $C(\kappa) = \lambda\kappa$, (77) yields

$$\kappa^* = \frac{1}{2}\log\frac{\sigma_0^2}{2\lambda}, \tag{78}$$

interior whenever $\sigma_0^2 > 2\lambda$.

LLR (Pomatto et al.) cost. With $C_{\text{LLR}}(\kappa) = a\,(e^{2\kappa} - 1)$, (77) yields

$$\kappa^* = \frac{1}{4}\log\frac{\sigma_0^2}{4a}, \tag{79}$$

[2] An alternative, dual route does not require positing a functional form for C at all. Behringer (2026) shows that under a *hard* capacity constraint $I(\Theta, S) \le \kappa$, the marginal cost of information is identified endogenously by convex (Fenchel) duality as the slope of the value of information, $\lambda^*(\kappa) = V'(\kappa)$, determined entirely by utility, prior, and action space. The convex penalty C assumed here is then a derived object rather than a primitive, the linear-penalty case corresponding to a (non-generic) linear V.

interior whenever $\sigma_0^2 > 4a$.

The linear RI cost yields $\kappa^* = \frac{1}{2}\log(\sigma_0^2/2\lambda)$ and the LLR cost $\kappa^* = \frac{1}{4}\log(\sigma_0^2/4a)$ share the same structure, with identical comparative statics.

Proposition 6 (Comparative Statics)**.** *Under either cost specification the optimal capacity rises with prior uncertainty, falls with the marginal cost of information, and is independent of the expected margin:*

1. $\dfrac{\partial\kappa^*}{\partial\sigma_0^2} > 0$ *(equal to* $\frac{1}{2\sigma_0^2}$ *under the linear cost and* $\frac{1}{4\sigma_0^2}$ *under the LLR cost): higher uncertainty increases information demand;*
2. $\dfrac{\partial\kappa^*}{\partial\lambda} = -\dfrac{1}{2\lambda} < 0$ *and* $\dfrac{\partial\kappa^*}{\partial a} = -\dfrac{1}{4a} < 0$*: a higher marginal cost of information reduces acquisition;*
3. $\dfrac{\partial\kappa^*}{\partial(\mu_0 - c)} = 0$*: the profit margin does not affect information acquisition (for quadratic profit).*

7. Two Information Sources: Allocating Attention Between Demand and Cost

We now let *both* the demand intercept and the marginal cost be unknown, and we give the firm two distinct informational channels, one for each. This is the natural industrial-organisation environment: demand information (market research, sales histories) and cost information (supplier quotes, accounting and engineering estimates) reach the firm through separate sources. The scarce resource is the total attention the firm can spread across them, and the economic question is how to *allocate* it.

7.1. Setup

Demand is $q = \theta - p$, but now *both* the intercept θ and the marginal cost c are uncertain, with independent Gaussian priors

$$\theta \sim \mathcal{N}(\mu_\theta, \sigma_\theta^2), \qquad c \sim \mathcal{N}(\mu_c, \sigma_c^2), \qquad \theta \perp c.$$

Profit is $\pi = (p-c)q = (\theta - q - c)\,q$. Writing the payoff-relevant state as the *markup base*

$$m \;\equiv\; \theta - c \sim \mathcal{N}(\mu_m, \sigma_m^2), \qquad \mu_m = \mu_\theta - \mu_c, \quad \sigma_m^2 = \sigma_\theta^2 + \sigma_c^2,$$

profit depends on (θ, c) only through m: the full-information quantity is $q^\star = m/2$ and full-information profit is $m^2/4$. The decision problem is thus one-dimensional in m, even though the *informational* problem is two-dimensional, because m can only be learned through its two components. (The single-dimension model is the special case $\sigma_c^2 \to 0$, cost known.)

Two channels. The firm observes

$$s_\theta = \theta + \varepsilon_\theta, \quad I(\Theta, S_\theta) = \kappa_\theta, \qquad s_c = c + \varepsilon_c, \quad I(C, S_c) = \kappa_c,$$

with noise variances fixed endogenously as in the generalisation of (29), $\sigma_{\varepsilon_i}^2 = \sigma_i^2/(e^{2\kappa_i} - 1)$ for $i \in \{\theta, c\}$. Because the priors and the two channels are independent, the total information processed is additive,

$$I\big((\Theta, C), (S_\theta, S_c)\big) = \kappa_\theta + \kappa_c, \tag{80}$$

so a budget on total mutual information is a budget on $\kappa_\theta + \kappa_c$. The firm splits a capacity κ across the two channels,

$$\kappa_\theta + \kappa_c \le \kappa, \qquad \kappa_\theta, \kappa_c \ge 0.$$

7.2. VoI with Two Sources

Given the two signals, certainty equivalence (profit is linear in m for fixed q) implies the firm sets $q = \mu_m(s)/2$ with $\mu_m(s) = E\{\theta \mid s_\theta\} - E\{c \mid s_c\}$, earning expected profit $\mu_m(s)^2/4$. As the channels are independent, the posterior variance of the markup base is the sum of the two channels' posterior variances,

$$\mathrm{Var}(m \mid s) = \sigma_\theta^2 e^{-2\kappa_\theta} + \sigma_c^2 e^{-2\kappa_c}.$$

Proposition 7 (Separable VoI)**.** *The VoI is additively separable across the two channels,*

$$\mathrm{VoI}(\kappa_\theta, \kappa_c) = \frac{\sigma_\theta^2}{4}\big(1 - e^{-2\kappa_\theta}\big) + \frac{\sigma_c^2}{4}\big(1 - e^{-2\kappa_c}\big), \tag{81}$$

each term being the single-channel value of (57) *evaluated at that channel's capacity.*

Proof. By the law of total variance, $\mathrm{Var}(\mu_m(s)) = \sigma_m^2 - \mathrm{Var}(m \mid s) = \sigma_\theta^2(1 - e^{-2\kappa_\theta}) + \sigma_c^2(1 - e^{-2\kappa_c})$. Ex-ante expected profit is $V = \frac{1}{4}\big[\,\mathrm{Var}(\mu_m(s)) + \mu_m^2\big]$, and subtracting the no-information value $V|_{\kappa_\theta=\kappa_c=0} = \mu_m^2/4$ gives (81). $\square$

7.3. Optimal Allocation: Reverse Water-Filling

The firm maximises (81) subject to $\kappa_\theta + \kappa_c = \kappa$; the constraint binds since VoI is strictly increasing in each argument.

Theorem 3 (Optimal attention allocation)**.** *Let* $\kappa > 0$*. The optimal allocation equalises the* posterior *(residual) variances across the two dimensions whenever an interior split is feasible,*

$$\sigma_\theta^2 e^{-2\kappa_\theta^\star} = \sigma_c^2 e^{-2\kappa_c^\star} \equiv \sigma^{\star 2}. \tag{82}$$

Explicitly, if $\left|\log(\sigma_\theta^2/\sigma_c^2)\right| \le 2\kappa$*,*

$$\kappa_\theta^\star = \frac{\kappa}{2} + \frac{1}{4}\log\frac{\sigma_\theta^2}{\sigma_c^2}, \qquad \kappa_c^\star = \frac{\kappa}{2} - \frac{1}{4}\log\frac{\sigma_\theta^2}{\sigma_c^2}. \tag{83}$$

If instead $\log(\sigma_\theta^2/\sigma_c^2) > 2\kappa$*, the firm specialises,* $\kappa_\theta^\star = \kappa$ *and* $\kappa_c^\star = 0$ *(and symmetrically when* $\log(\sigma_c^2/\sigma_\theta^2) > 2\kappa$*).*

Proof. The objective is concave and the constraint linear. The interior first-order conditions equate the marginal values of attention, $\frac{\sigma_\theta^2}{2}e^{-2\kappa_\theta} = \frac{\sigma_c^2}{2}e^{-2\kappa_c} = \nu$, which is (82); solving with $\kappa_\theta + \kappa_c = \kappa$ yields (83). Feasibility of $\kappa_c^\star \ge 0$ is exactly $\log(\sigma_\theta^2/\sigma_c^2) \le 2\kappa$. Otherwise the marginal value of the first bit sent to the low-variance channel, $\sigma_c^2/2$, lies below the marginal value of every bit sent to the high-variance channel, so the solution is at the corner. $\square$

Equation (82) is the *reverse water-filling* of rate-distortion theory (see Cover and Thomas (2006) 10.3.3): the firm splits its budget evenly and then *tilts* attention towards the more uncertain dimension by a quarter of the log variance-ratio. When the asymmetry in prior uncertainty is large relative to the budget, the firm ignores the more predictable dimension altogether and learns only the volatile one, a low-variance source receives *no* attention until the budget is large enough to make its first bit worthwhile.

Corollary 3 (Value at the optimum)**.** *In the interior regime the common posterior variance is the geometric-mean benchmark $\sigma_\theta \sigma_c \, e^{-\kappa}$, and*

$$\mathrm{VoI}^\star(\kappa) = \frac{1}{4}\Big(\sigma_\theta^2 + \sigma_c^2 - 2\,\sigma_\theta\sigma_c\, e^{-\kappa}\Big). \tag{84}$$

As $\kappa \to \infty$ this rises to $\frac{1}{4}(\sigma_\theta^2 + \sigma_c^2) = \sigma_m^2/4$, the full-information value of Marschak (1954).

Proof. Multiplying the two sides of (82) (note that the budget binds) gives $\sigma_\theta^2 \sigma_c^2 e^{-2\kappa} = (\sigma^\star)^4$, so the common posterior variance is $\sigma^{\star 2} = \sigma_\theta \sigma_c e^{-\kappa}$. Substituting into $\mathrm{VoI} = \frac{1}{4}[\sigma_\theta^2 + \sigma_c^2 - 2\sigma^{\star 2}]$ yields (84). □

Proposition 8 (Comparative statics of the allocation)**.** *In the interior regime,*

$$\frac{\partial \kappa_\theta^\star}{\partial \sigma_\theta^2} = \frac{1}{4\sigma_\theta^2} > 0, \qquad \frac{\partial \kappa_\theta^\star}{\partial \sigma_c^2} = -\frac{1}{4\sigma_c^2} < 0, \qquad \frac{\partial \kappa_\theta^\star}{\partial \kappa} = \frac{1}{2},$$

and the split is independent of the prior means (μ_θ, μ_c), hence of the expected margin $\mu_\theta - \mu_c$.

The firm therefore attends more to whichever primitive is more uncertain, shifts attention *away* from a dimension as the other becomes more volatile, and divides any marginal capacity equally. As in the one-dimensional model, the profit margin is irrelevant to the information decision under quadratic profit; here it is the *relative* uncertainty of demand and cost, not their levels or the markup, that governs the allocation.

7.4. Endogenous Total Capacity

If the firm also chooses how much total capacity to acquire at linear cost $C = \lambda(\kappa_\theta + \kappa_c)$, the two channels *decouple*: each first-order condition $\frac{\sigma_i^2}{2} e^{-2\kappa_i} = \lambda$ involves only its own capacity, giving

$$\kappa_\theta^\star = \frac{1}{2}\log\frac{\sigma_\theta^2}{2\lambda}, \qquad \kappa_c^\star = \frac{1}{2}\log\frac{\sigma_c^2}{2\lambda}, \tag{85}$$

each active iff its prior variance exceeds the threshold 2λ. The induced total capacity,

$$\kappa^\star = \kappa_\theta^\star + \kappa_c^\star = \log \frac{\sigma_\theta \sigma_c}{2\lambda}, \tag{86}$$

is governed by the *geometric mean* of the two uncertainties, while the tilt $\kappa_\theta^\star - \kappa_c^\star = \frac{1}{2}\log(\sigma_\theta^2/\sigma_c^2)$ coincides with the fixed-budget water-filling of Theorem 3. The fixed-budget and cost-based formulations thus agree on *how* attention is split and differ only on *how much* is acquired, the budget and price formulations of the allocation thus coinciding on the split.

7.5. Industrial-Organisation Implications

The reduction to a single payoff-relevant state $m = \theta - c$ keeps the allocation tractable, yet the allocation is genuinely two-dimensional and delivers cross-sectional predictions absent from the one-signal model. Firms in volatile-demand environments devote more attention to demand and misprice more on cost; firms facing volatile inputs do the reverse; and firms whose uncertainty is concentrated in one dimension specialise, leaving the predictable dimension at its prior. Because the split depends on the ratio $\sigma_\theta^2/\sigma_c^2$ and not on margins, two equally profitable firms with different uncertainty profiles allocate attention, and therefore err in pricing, in systematically different ways. These predictions presume the firm gathers demand and cost information through separate channels; whether that separation is itself optimal is the question we turn to next.

7.6. Relation to Unrestricted Rational Inattention

Siloing is a restriction relative to unrestricted RI (Sims, 2003; Matějka and McKay, 2015), in which the firm designs the joint information structure freely subject only to a budget on total mutual information. Since profit depends on (θ, c) only through the markup base $m = \theta - c$, a firm free to choose what to learn never resolves the two primitives separately.

Proposition 9 (Payoff-relevant sufficiency)**.** *For any signal S with $I((\Theta, C), S) \leq \kappa$, it is optimal to learn m alone: the optimal signal is*

$S = m + \varepsilon$ *with* $I(M, S) = \kappa$, *giving* $\mathrm{Var}(m \mid S) = \sigma_m^2 e^{-2\kappa}$ *and*

$$\mathrm{VoI}_{\mathrm{RI}}(\kappa) = \frac{\sigma_m^2}{4}\big(1 - e^{-2\kappa}\big). \tag{87}$$

Proof. Profit loss is proportional to $\mathrm{Var}(m \mid S)$. As m is a function of (θ, c), the data-processing inequality gives $I(M, S) \le \kappa$, and the Gaussian maximum-entropy bound (Cover and Thomas, 2006, Thm. 8.6.5) gives $\mathrm{Var}(m \mid S) \ge \sigma_m^2 e^{-2\kappa}$, attained by $S = m + \varepsilon$ with Gaussian ε. $\square$

The two-dimensional problem thus collapses *endogenously* to the one-dimensional model (57) with $\sigma_m^2 = \sigma_\theta^2 + \sigma_c^2$, and needs no independence assumption, since m is payoff-relevant whatever the correlation. The demand–cost allocation of Theorem 3 is meaningful only because the siloed technology forbids the integrated signal that pure inattention would select. Because every pair of dedicated channels is one feasible structure for the unrestricted firm, $\mathrm{VoI}_{\mathrm{RI}} \ge \mathrm{VoI}^\star$, with interior gap $\frac{1}{4}e^{-\kappa}\big(2\sigma_\theta\sigma_c - (\sigma_\theta^2 + \sigma_c^2)e^{-\kappa}\big) > 0$ for finite κ. This gap is the *value of informational integration*: the profit forgone by gathering demand and cost information in separate silos rather than fusing them into the payoff-relevant statistic, an organisational-design object with no counterpart in the single-dimension model.

The Gaussian signal about m is special to the linear-quadratic structure. Were the action set discrete, the same logic would instead deliver the multinomial-logit choice probabilities and discrete pricing of Matějka and McKay (2015), the two being faces of a common RI foundation, selected by whether pricing is continuous or discrete.

7.7. Correlated Demand and Cost

The independence of θ and c is what makes the two channels complementary and the allocation a clean water-filling. Allowing joint normality with correlation ρ, the markup base has prior variance $\sigma_m^2 = \sigma_\theta^2 + \sigma_c^2 - 2\rho\,\sigma_\theta\sigma_c$, decreasing in ρ: when demand and cost move together, the margin $\theta - c$ is intrinsically more stable, so there is less to learn.

The unrestricted benchmark is unchanged in form, since m remains the payoff-relevant statistic and Proposition 9 still gives $\mathrm{VoI}_{\mathrm{RI}}(\kappa) = \frac{\sigma_m^2}{4}(1 - e^{-2\kappa})$ with the correlation folded into σ_m^2. The siloed problem, by contrast, loses its additive structure: a signal about θ is now also informative about c, so the two channels become *substitutes* rather than complements. The firm cross-infers at the updating stage, the joint information of the two channels is no longer $\kappa_\theta + \kappa_c$, and the closed-form split of Theorem 3 gives way to a coupled condition that is generically numerical. Qualitatively, positive correlation renders one channel partly redundant and pushes the firm to *concentrate* attention rather than diversify it, while the value of informational integration is partly recovered through cross-inference and shrinks as $|\rho| \to 1$. This parallels the correlated-error extension in Section 3.3, correlation across fundamentals here, across signal noise there.

8. Conclusion

We have developed a mutual- and directed-information framework for monopoly pricing under demand uncertainty that clarifies the non-monotonicities reported by Behringer (2021). An exact mutual information formula maps Behringer's additive-multiplicative noise model into an information capacity κ; parameterised by κ rather than by the noise variances, the VoI is monotone and concave, in closed form, for Gaussian priors in both static and dynamic settings, so the apparent non-monotonicity is an artefact of the noise parameterisation rather than of learning itself. Directed information supplies the causal accounting for the multi-period problem, in which prices shape the signals that drive subsequent beliefs.

Letting the firm face uncertainty in both demand and marginal cost, each learned through a separate channel, the framework yields an attention-allocation theory of where firms optimally misprice: the optimal split follows a reverse water-filling rule governed by the ratio of demand to cost uncertainty and independent of the expected margin, with the single-dimension model as its special case. Confronting this siloed firm with the unrestricted Rational Inattention benchmark shows the allocation to be meaningful only because separate channels forbid the integrated signal that pure inattention would select; the resulting profit gap measures the *value of informational integration*, an organisational-design object absent from the one-dimensional model.

These closed-form value functions connect naturally to the risk-measure construction of Behringer (2026): the Value of Information (VoI) studied here is the object from which the convex risk measures of that framework are derived, and the dynamic extension points towards their intertemporal counterparts. The analysis sits within the programme of Marschak (1954), who first sought to bring information theory into the core of economic decision-making. More broadly, the VoI analysis offers a transparent foundation for studying information acquisition in markets with demand uncertainty and endogenous pricing.

References

Behringer, S. (2021). Multiplicative Normal Noise and Nonconcavity in the Value of Information. *Theoretical Economics Letters*, 11, 116–124. https://www.scirp.org/pdf/tel_2021022410371431.pdf

Behringer, S. and Belavkin, R. V. (2023). The Value of Information and Circular Settings. *arXiv preprint* https://arxiv.org/abs/2303.16126

Behringer, S. and Belavkin, R. V. (2025). The Value of Information in Economic Contexts. *Physical Sciences Forum*, 12(1), 6. https://doi.org/10.3390/psf2025012006

Behringer, S. (2026). Hard Constraint, Convex Duality, and the Endogenous Cost of Information. mimeo, available at https://www.stefanbehringer.com.

Cover, T. M. and Thomas, J. A. (2006). *Elements of Information Theory*, 2nd edition. Wiley-Interscience.

Denti, T. (2022). Posterior Separable Cost of Information. *American Economic Review*, 112(10): 3215–3259. doi:10.1257/aer.20211252.

Grossman, S. J. and Stiglitz, J. E. (1980). On the Impossibility of Informationally Efficient Markets. *American Economic Review*, 70(3): 393–408.

Hellwig, M. F. (1980). On the Aggregation of Information in Competitive Markets. *Journal of Economic Theory*, 22(3): 477–498.

Marschak, J. (1954). Towards an Economic Theory of Organization and Information. In R. M. Thrall, C. H. Coombs, and R. L. Davis (eds.), *Decision Processes*, pp. 187–220. Wiley, New York.

Massey, J. L. (1990). Causality, Feedback and Directed Information. *Proceedings of the International Symposium on Information Theory and Its Applications (ISITA)*, Waikiki, Hawaii. http://www.isiweb.ee.ethz.ch/archive/massey_pub/pdf/BI532.pdf

Matějka, F. and McKay, A. (2015). Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model. *American Economic Review*, 105(1): 272–298. `https://doi.org/10.1257/aer.20130047`

Pomatto, L., Strack, P., and Tamuz, O. (2023). The Cost of Information: The Case of Constant Marginal Costs. *American Economic Review*, 113(5): 1360–1393. doi:10.1257/aer.20190185.

Sims, C. A. (2003). Implications of Rational Inattention. *Journal of Monetary Economics*, 50(3): 665–690. `https://doi.org/10.1016/S0304-3932(03)00029-1`

Stratonovich, R. L. (1975/2020). *Theory of Information and Its Value*. Translated and edited by Roman V. Belavkin, Panos M. Pardalos, and Jose C. Principe. Springer, Cham. ISBN: 978-3-030-22832-3.

Whitmeyer, M. (2025). A concavity in the value of information. *Games and Economic Behavior*, **154**, 200–207. `https://doi.org/10.1016/j.geb.2025.09.003`