Submitted to *Econometrica*

HARD CONSTRAINTS, CONVEX DUALITY, AND THE ENDOGENOUS COST OF INFORMATION

STEFAN BEHRINGER[1]
Value of Information Group, Universität Bielefeld

The rational inattention literature employs two formulations of information costs: a hard constraint bounding mutual information by a Shannon capacity, and a soft constraint penalising mutual information by a linear entropy penalty. We establish the structural relationship between these formulations via convex duality theory and develop its consequences. First, the hard-constraint value function, the Stratonovich/Shannon Value of Information (VoI), is the primitive object of rate distortion theory. The soft-constraint objective is its Fenchel conjugate, recoverable without parametric restriction, and the marginal cost of information is identified endogenously as the slope of the VoI; the linear entropy penalty is therefore a special case, not a maintained assumption. Second, the VoI generically depends on the full posterior distribution; it collapses to a function of posterior variance alone if and only if utility is quadratic, the prior is Gaussian, and optimal signals preserve conjugacy, three jointly necessary conditions. Third, the optimal signal distribution takes an exponential form and generates a class of convex risk measures; in a portfolio application, tail risk substantially amplifies information value relative to a mean-variance equivalent benchmark, a distinction that variance-based analysis cannot detect.

KEYWORDS: rational inattention, Stratonovich/Shannon Value of Information (VoI), hard and soft information constraints, convex duality, Fenchel conjugate, endogenous information costs, Gaussian quadratic benchmark, convex risk measures, tail risk.

1. INTRODUCTION

The rational inattention (RI) model, introduced by Sims (2003), captures the idea that attention is a scarce resource. Agents optimally allocate their limited information processing capacity, measured in bits, across the dimensions of uncertainty they face. Two formulations have emerged in the literature. The *hard-constraint* formulation, as in Sims (2003), bounds mutual information by an exogenous Shannon capacity $\bar{I}$. The *soft-constraint* formulation, predominant in the subsequent literature surveyed by Maćkowiak, Matějka, and Wiederholt (2023), penalises mutual information by a linear entropy cost $\lambda I(\Theta, S)$, where $\lambda > 0$ is an exogenously imposed parameter.

Azrieli (2021) emphasizes that the two formulations are locally equivalent, yielding identical solutions for a fixed value of the Lagrange multiplier via standard Lagrangian duality provided the capacity constraint binds, but globally non-equivalent, as the multiplier varies endogenously across decision problems and can generate comparative statics reversals. The structural relationship between the two formulations has not, however, been grounded in convex duality theory, nor have the precise conditions under which the linear entropy penalty is a valid maintained assumption been formally established.

This paper fills that gap and makes three main contributions.

The first contribution is to establish the convex duality structure of the RI problem. We show that $V(\bar{I})$ is the information-utility frontier of rate distortion theory in the sense of Stratonovich

Stefan Behringer[1] : stefan@stefanbehringer.com

[1]The author is grateful to Roman V. Belavkin, Julien Berguinot, Michel de Lara, Valérie Girardin, Terry Rockafellar, and Laura Veldkamp, to seminar participants at GT EMS: Groupe de Travail Entropie (Université de Caen), ENPC Paris, and the DIS Conference (London, 2025), and to a reviewer for comments that motivated some results of this paper.

(1975/2020), and that the soft-constraint objective is its Fenchel conjugate (Theorem 1). By the biconjugate theorem (Theorem 2), $V(\bar{I})$ is fully recoverable from from the soft-constraint objective without any parametric assumption on the cost functional form. By the inverse subgradient theorem (Theorem 3), the marginal cost of information is identified endogenously from the curvature of V, determined entirely by the utility function, the prior, and the action space. The linear penalty of the soft-constraint formulation is therefore not a maintained assumption but a special case, valid only when $V(\bar{I})$ is linear in $\bar{I}$, a non-generic condition. These results provide the formal foundation for Sims's (2003, 2006) advocacy of the hard-constraint approach, particularly in financial markets where information costs are decision-problem specific and generically non-linear in bits.

The second contribution is to characterise precisely the conditions under which the Gaussian quadratic benchmark, called the linear-quadratic case by Sims (2006), which underlies the canonical closed-form solutions of rate distortion theory and the RI literature and appears independently in the Kalman filter and mean-variance portfolio theory, is valid. Theorem 4 establishes a three-way equivalence: information value depends only on posterior variance if and only if utility is quadratic, the prior is Gaussian, and optimal signals preserve conjugacy. Each of these three conditions is necessary, not merely sufficient; removing any one of them causes the full distribution of posterior beliefs to enter the optimality conditions, not merely its first two moments. Any theory that makes use of the Gaussian quadratic benchmark implicitly imposes all three conditions simultaneously.

The third contribution is to develop the general Stratonovich/Shannon Value of Information framework which is rooted in classical rate distortion theory (Stratonovich, 1975/2020, Behringer and Belavkin, 2023, 2025). The optimal signal distribution takes an exponential form (Proposition 3), with the Lagrange multiplier on the capacity constraint playing the role of a shadow price of information, sometimes called an *information temperature*. The certainty equivalent of the signal lottery available in each state, called the *free energy* of that state, aggregates utility over signals in that state, and the primal-dual identity (Lemma 2) recovers $V(\bar{I})$ as the prior expectation of these state-level free energies plus the total shadow cost of capacity. The resulting value function generates a class of convex risk measures (Theorem 5) satisfying extended coherence in the sense of Rockafellar (2007) but failing positive homogeneity, a structural consequence of the non-homogeneity of Shannon mutual information with distinct empirical implications for scale effects in information acquisition.

Our work contributes to a methodological strand that brings risk measure theory to bear on economic problems (Gershkov, Moldovanu, Strack, and Zhang, 2025) by deriving a risk measure from first principles of information acquisition under hard capacity constraints, so that the value function is an output of the problem rather than an exogenous primitive.

The paper connects several literatures that have developed largely independently. Rational inattention originated with hard information constraints in Sims (2003). Subsequent models (Van Nieuwerburgh and Veldkamp, 2010, Matějka and McKay, 2015) typically use soft constraints with costs proportional to mutual information. Although hard and soft formulations are locally dual at interior optima, they generate distinct comparative statics globally, as formalised by Denti, Marinacci, and Montrucchio (2020). Robust control approaches (Hansen and Sargent, 2001, 2008) share the entropy-based structure but optimise over alternative probability models rather than over signal distributions. Convex risk measure theory (Föllmer and Schied, 2004, Rockafellar, 2007) treats information structures as exogenous; hard information constraints bridge these literatures while generating predictions absent from each individually. De Lara and Gossner (2020) exploit the duality between actions and beliefs to characterise the value function via convex analysis and derive conditions for positive value of information from local properties at the prior; their approach is complementary to the duality results in Section 3.

Denti, Marinacci, and Rustichini (2022) show that posterior-based cost representations, which include the standard entropy penalty of the soft-constraint formulation, are in many cases inconsistent with any primitive model of costly experimentation; the hard-constraint formulation is immune to this critique since it imposes a bound directly on the experiment rather than on the distribution of posterior beliefs.

A related strand of the literature pursues a different strategy: rather than taking Shannon mutual information as the cost functional, it seeks to enrich the family of admissible cost functions. De Oliveira, Denti, Mihm, and Ozbek (2017) show how hidden information costs can be identified from menu-choice data under three canonical properties of the cost function; their approach takes the soft-side cost functional as the object to be recovered. Fosgerau, Melo, de Palma, and Shum (2020) extend the Shannon-logit equivalence to the entire class of additive random utility discrete choice models by replacing Shannon entropy with a generalised entropy, with information costs taking the form of Bregman divergences. Bloedel, Denti, and Pomatto (2025) introduce a cost function based on multivariate f-divergences that generalises both Shannon mutual information and the class of posterior-separable cost functions, analysing it via convex duality and Arrow-Pratt methods.

All of these contributions work on the soft side of the hard-soft distinction: they ask which cost functional form to impose or recover, taking the cost function as a primitive of the model. The enrichment of the cost class, from Shannon entropy to Bregman divergences or f-divergences, extends the descriptive reach of the soft approach but does not alter its epistemological structure: the cost functional remains an input chosen by the researcher rather than an output of the decision problem. The hard-constraint approach inverts this structure. By Theorems 2 and 3, the information constraint is the primitive, and the implied cost schedule determined endogenously from the curvature of VoI, itself an output of the utility function, the prior, and the action space. This distinction is most consequential in the settings where richer cost classes are most needed: those where information costs are genuinely unknown and decision-problem specific, as in the financial market and climate applications developed below.

Section 2 defines the two formulations and develops the framework. Section 3 states the main duality results. Section 4 analyses the Gaussian quadratic benchmark and its knife-edge character. Section 5 develops the Gibbs framework. Section 6 defines the VoI risk measure. Section 7 presents applications. Section 8 concludes. Proofs are collected in the Appendix.

2. FRAMEWORK

2.1. *Basic Setup*

Let Θ denote the state of the world with prior distribution $\mu \in \Delta(\Omega)$ over a finite state space Ω, and let $\mathcal{A}$ denote the finite action space. The agent chooses a signal S via the conditional distribution $p(s|\theta)$ and selects the Bayesian optimal action

$$a^*(s) = \arg\max_{a \in \mathcal{A}} E\{U(a,\Theta) \mid S = s\},$$

where $U : \mathcal{A} \times \Omega \to \mathbb{R}$ is the utility function. Shannon mutual information between Θ and S is

$$I(\Theta, S) = \int p(\theta, s) \log \frac{p(s|\theta)}{p(s)} \, d\theta \, ds, \tag{1}$$

measured in bits, where $p(\theta, s) = \mu(\theta) p(s|\theta)$ is the joint distribution and $p(s) = \int \mu(\theta) p(s|\theta) \, d\theta$ is the marginal signal distribution.

2.2. *The Stratonovich/Shannon Value of Information*

The central object of this paper is the Stratonovich/Shannon Value of Information, defined as the supremum of expected utility subject to a mutual information constraint:

$$U(\bar{I}) \equiv \sup_{p(s|\theta)} \left[E_\theta\{U(a^*(S),\theta)\} \text{ subject to } I(\Theta,S) \le \bar{I} \right]. \tag{2}$$

The Stratonovich/Shannon Value of Information (*VoI*) at information budget $\bar{I}$ is then

$$V(\bar{I}) = U(\bar{I}) - U(0), \tag{3}$$

where $U(0) = \max_{a\in\mathcal{A}} E\{U(a,\Theta)\}$ is the expected utility achieved by the prior-optimal action with no signal. Since $U(0)$ depends only on μ and U and not on $\bar{I}$, it is a constant with respect to the optimisation and drops out of all derivative-based results, in particular $V'(\bar{I}) = U'(\bar{I})$.

The function $V(\bar{I})$ is the information-utility frontier of rate distortion theory in the sense of Stratonovich (1975/2020), playing the same role in information economics that the efficient frontier plays in portfolio theory: the primitive object from which all optimal tradeoffs are derived.[2]

The certainty equivalent will play a central role in characterising the optimal signal distribution and the value function $U(\bar{I})$ in Section 5. For a random variable X and information price $\lambda > 0$, define

$$\mathrm{CE}_\lambda(X) = \lambda \log E\left\{\exp\left(\frac{X}{\lambda}\right)\right\}. \tag{4}$$

As $\lambda\to\infty$, $\mathrm{CE}_\lambda(X)\to E\{X\}$ (risk neutrality); as $\lambda\to 0$, $\mathrm{CE}_\lambda(X)\to\max X$ (maximin behaviour). The certainty equivalent satisfies translation invariance: $\mathrm{CE}_\lambda(X+c) = \mathrm{CE}_\lambda(X)+c$ for constants c.

2.3. *The Two Formulations and Local Equivalence*

DEFINITION 1—Hard and Soft Information Constraints:
1. Hard constraint (Shannon capacity):

$$V(\bar{I}) = \max_{p(s|\theta)} E\{U(a^*(S),\Theta)\} \quad \textit{subject to} \quad I(\Theta,S)\le\bar{I}, \tag{5}$$

where $\bar{I} > 0$ is an exogenous information capacity bound.
2. Soft constraint (entropy penalty):

$$V_{\text{soft}}(\lambda) = \max_{p(s|\theta)} \left\{E\{U(a^*(S),\Theta)\} - \lambda I(\Theta,S)\right\}, \tag{6}$$

where $\lambda > 0$ is an exogenously imposed penalty parameter.

[2] Stratonovich (1975/2020) (p. 301, eq. (9.3.7)) calls $V(I) = R_0 - R_+(I)$ the *value of Shannon's information*, where $R_+(I)$ is the minimum expected cost subject to the information constraint and R_0 is the prior-optimal baseline cost. Since Stratonovich works with expected costs rather than utilities, $R = -U$, and $R_+(I)\le R_0$ so that $V(I)\ge 0$. Our formulation (3) is the utility-maximisation counterpart. Throughout this paper we use λ for the Lagrange multiplier on the capacity constraint, following the conventions of the rational inattention literature; Stratonovich (1975/2020) uses β for the same quantity, following the statistical mechanics convention in which $\beta = 1/(k_\mathrm{B}T)$ denotes inverse temperature.

The Lagrangian of (5) is $\mathcal{L} = E\{U(a^*(S), \Theta)\} - \lambda[I(\Theta, S) - \bar{I}]$, which is structurally identical to (6) for the same λ. The first-order conditions therefore coincide for any fixed λ. By KKT complementary slackness, $\lambda \cdot [I(\Theta, S) - \bar{I}] = 0$ with $\lambda > 0$ implies the constraint binds at the optimum.[3]

PROPOSITION 1—Local Equivalence: *When the capacity constraint binds, the two formulations yield identical solutions for $\lambda^* = V'(\bar{I})$. The multiplier λ^* is determined endogenously in the hard formulation and must be imposed exogenously in the soft formulation.*

REMARK: *Local equivalence rests on the linearity of $\lambda I(\Theta, S)$ in mutual information. For a general cost $c(I)$, the Lagrangian of (5) does not coincide with the soft-constraint objective and the equivalence breaks down. The hard-constraint formulation accommodates non-linear cost schedules without modification, since it imposes no functional form on costs at all.*

2.4. *Regularity*

The following regularity conditions are maintained throughout the paper.

ASSUMPTION 1:

(R1) *Finite second moments. The prior μ has finite second moments, so that the posterior variance is well defined for all feasible signal structures and the comparison across signal distributions in the main results is meaningful.*

(R2) *Finite mutual information. The Shannon mutual information $I(\Theta, S)$ is finite for all signal distributions under consideration. With finite state and action spaces this is satisfied for all signal distributions that do not perfectly reveal the state, but is stated explicitly to cover the case of a continuous signal space.*

(R3) *Interior solution with the constraint binding. The information budget $\bar{I}$ satisfies $0 < \bar{I} < C$, where C is the channel capacity, so that the constraint binds at the interior optimum. At boundary values the capacity constraint may not bind: at $\bar{I} = 0$ information has zero value, and at $\bar{I} \geq C$ the constraint is slack.*

Under Assumption 1, the VoI $V(\bar{I})$ is continuous, strictly increasing, and strictly concave on $(0, C)$, reflecting diminishing returns to information: the first bit of information is more valuable than subsequent bits. The subdifferential $\partial V(\bar{I})$ is non-empty at every interior point.

3. CONVEX DUALITY AND ENDOGENOUS INFORMATION COSTS

3.1. *The Fenchel Conjugate Relationship*

THEOREM 1—Fenchel–Legendre Representation: *Under Assumption 1, the soft-constraint value function admits the representation*

$$V_{\text{soft}}(\lambda) = \sup_{\bar{I} \geq 0} \{V(\bar{I}) - \lambda \bar{I}\}.$$

That is, V_{soft} is the concave Fenchel–Legendre transform of the VoI $V(\bar{I})$.

[3]Global non-equivalence, arising from the endogenous variation of λ^* across decision problems, is established by Azrieli (2021)

Theorem 1 identifies the soft-constraint objective as the Fenchel–Legendre conjugate of the VoI with respect to λ. The transform is invertible when the VoI is strictly concave, which holds under Assumption 1 whenever the decision problem is non-trivial, establishing a one-to-one correspondence between information budgets $\bar{I}$ and shadow prices λ^* on the interior of the domain.

3.2. *Biconjugate Recovery Without Parametric Restriction*

THEOREM 2—Biconjugate Recovery:

$$V(\bar{I}) = \inf_{\lambda \geq 0} \{ V_{\text{soft}}(\lambda) + \lambda \bar{I} \}.$$

The VoI $V(\bar{I})$ is fully recoverable from $V_{\text{soft}}(\lambda)$ without any parametric assumption on the cost functional form.

Since $V(\bar{I})$ is continuous and concave under Assumption 1, $-V(\bar{I})$ is closed and convex. Theorem 2 is the biconjugate identity applied to $-V$ (Rockafellar, 1974, Theorem 5, p. 16): $(-V)^{**} = \mathrm{cl}\,\mathrm{co}(-V) = -V$, where $\mathrm{co} f$ denotes the convex envelope and cl its closure, both of which leave $-V$ unchanged. The Fenchel transform therefore induces a one-to-one correspondence between the hard-constraint value function and the soft-constraint objective, with no restriction on the functional form of information costs.

3.3. *Endogenous Identification of the Marginal Cost*

THEOREM 3—Endogenous Marginal Cost: *At any $\bar{I}$ satisfying Assumption 1,*

$$\partial V(\bar{I}) = \{ \lambda^* \mid \bar{I} \in \partial V_{\text{soft}}(\lambda^*) \}.$$

When V is differentiable, this collapses to $\lambda^*(\bar{I}) = V'(\bar{I})$: primal and dual solutions identify each other mutually.

Theorem 3 is the inverse subgradient identity applied to $-V$ (Rockafellar, 1974, Theorem 12, p. 35). It states that λ^* is a subgradient of V at $\bar{I}$ if and only if $\bar{I}$ is a subgradient of the conjugate V_{soft} at λ^*, so that neither object is more primitive than the other: the hard-constraint value function and the soft-constraint objective are dual representations of the same information-utility frontier.

PROPOSITION 2—Endogenous Non-Linear Information Costs: *The marginal cost $\lambda^*(\bar{I}) = V'(\bar{I})$ is determined entirely by U, μ, and $\mathcal{A}$. Since the VoI $V(\bar{I})$ is generically non-linear in $\bar{I}$, the implied cost schedule is generically non-constant. The linear penalty $\lambda I(\Theta, S)$ is a special case arising only when $V(\bar{I})$ is linear in $\bar{I}$, a non-generic condition.*

Proposition 2 supplies the formal basis for Sims's (2003, 2006) advocacy of the hard-constraint approach. In financial markets, information costs are dominated by *costly investigation*, the hard-constraint cost of acquiring non-public information through experimentation, surveying, or proprietary research rather than by *Shannon capacity costs*, the soft-constraint cost of mapping freely available information into action given limited processing capacity. Investigation costs are not well measured in bits at a constant price: a single bit about a relevant asset may require extremely costly proprietary research while many irrelevant bits are nearly free. The cost per bit is therefore decision-problem specific and generically non-linear in bits, making the linear penalty $\lambda I(\Theta, S)$ inappropriate as a maintained assumption. The hard-constraint formulation avoids this difficulty: $\lambda^*(\bar{I}) = V'(\bar{I})$ is derived endogenously from the curvature of the VoI $V(\bar{I})$ with no parametric assumption required.

The shadow price $\lambda^*(\bar{I}) = V'(\bar{I})$ is also the maximum the agent would pay for a marginal unit of additional information capacity, well-defined and directly interpretable in terms of decision problem primitives. The soft-constraint formulation offers no such foundation: λ is imposed exogenously and no condition within the problem pins it down.

4. THE GAUSSIAN QUADRATIC BENCHMARK

4.1. *Setup and Closed-Form Value of Information*

The Gaussian quadratic case, called the linear-quadratic case by Sims (2006)[4], is the canonical example of rate distortion theory precisely because it is one of the rare cases admitting a closed-form rate-distortion function (Shannon, 1948, Sims, 2003, Cover and Thomas, 2006). An agent chooses action $a \in \mathbb{R}$ after observing signal S about an uncertain state $\Theta \sim N(\mu_0, \sigma_\theta^2)$. The signal structure is $S = \Theta + \varepsilon$ with $\varepsilon \sim N(0, \sigma_\varepsilon^2)$ independent of Θ. Utility is quadratic: $U(a,\theta) = -(a-\theta)^2$. The Bayesian optimal action equals the posterior mean, and for jointly Gaussian random variables the posterior variance satisfies $\mathrm{Var}(\Theta|S) = \sigma_\theta^2 e^{-2I}$, yielding expected utility $U(I) = -\sigma_\theta^2 e^{-2I}$.

LEMMA 1—Gaussian Quadratic VoI: *Under Gaussian quadratic structure with Gaussian signals, the VoI depends only on mutual information:*

$$V(I) = \sigma_\theta^2 \left(1 - e^{-2I}\right). \tag{7}$$

This function is strictly increasing, strictly concave, satisfies $V(0) = 0$, *and* $\lim_{I\to\infty} V(I) = \sigma_\theta^2$. *The marginal cost of information is* $\lambda^*(I) = V'(I) = 2\sigma_\theta^2 e^{-2I}$, *strictly decreasing in* I.

Any two signals with the same mutual information yield identical value in the Gaussian quadratic case. This sufficiency of mutual information for the VoI does not hold in general: outside the knife-edge conditions identified below, the full signal distribution matters.

4.2. *The Knife-Edge Sufficiency Conditions*

The tractability of the Gaussian quadratic case rests on three conditions that can fail independently. We now characterise exactly when they must hold simultaneously.

DEFINITION 2—Gaussian Quadratic Structure: *A decision problem with prior* μ *and utility* $U(a,\theta)$ *exhibits* Gaussian quadratic structure *if:*

(G1) ***Conjugacy.*** *Signals are Gaussian so that posterior beliefs remain Gaussian.*

(G2) ***State independence.*** *The normalising constant in the optimal signal distribution does not depend on the state* θ.

(G3) ***Mean-variance sufficiency.*** *Optimal actions and expected utility depend only on the first two moments of the posterior distribution.*

[4] In the Gaussian quadratic setting the optimal signal is a linear function of the state (the Kalman filter is the optimal linear filter under Gaussianity), and the optimal action is a linear function of the signal (the posterior mean is linear in the observation under Gaussian conjugacy).

THEOREM 4—Characterisation of Mean-Variance Sufficiency: *Consider a decision problem with prior* μ*, utility* $U(a,\theta)$*, and a hard information constraint as in* (1)–(2)*, subject to Assumption 1. The following statements are equivalent:*

(i) *The problem exhibits Gaussian quadratic structure in the sense of Definition 2.*
(ii) $U(a,\theta) = -(a-\theta)^2$ *up to positive affine transformation, the prior* μ *is Gaussian, and optimal signals are Gaussian.*
(iii) *The VoI under the hard Shannon mutual information constraint depends only on the posterior variance reduction, uniformly over all feasible signal distributions and all information budgets* $\bar{I} \in (0, C)$.

Moreover, if any one of (G1)–(G3) *fails, then generically the VoI depends on posterior moments beyond variance.*

Theorem 4 makes precise the sense in which Gaussian quadratic assumptions matter: each of the three conditions is necessary, not merely convenient.[5] Removing any one of them means that skewness, kurtosis, and other higher moments of the posterior enter the VoI, effects that a posterior-variance-based characterisation cannot capture. Any theory that makes use of the Gaussian quadratic benchmark implicitly imposes all three conditions simultaneously.

5. THE STATISTICAL MECHANICAL STRATONOVICH/SHANNON VOI FRAMEWORK

5.1. *Lagrangian Duality and the Optimal Signal*

The hard-constraint problem (5) is solved via its Lagrangian relaxation. Introducing multiplier $\lambda \geq 0$ on the capacity constraint yields

$$\mathcal{L}[p(s|\theta)] = E\{U(a^*(S), \Theta)\} - \lambda I(\Theta, S). \tag{8}$$

The variational first-order condition characterises the optimal signal distribution. For corner solutions where the capacity constraint does not bind, one sets $\lambda = 0$. In general, λ^* is found by bisection on the strictly monotone residual $I(\lambda) - \bar{I}$, which converges to arbitrary precision.

PROPOSITION 3—Optimal Signal Distribution: *At an interior solution to* (5)*, the optimal signal distribution satisfies*

$$p^*(s|\theta) = \frac{p(s)}{Z(\theta)} \exp\left(\frac{U(a^*(s),\theta)}{\lambda}\right), \tag{9}$$

where $\lambda^* = V'(\bar{I})$ *satisfies* $I(\Theta, S) = \bar{I}$ *and the partition function*[6] *is*

$$Z(\theta) = \int p(s) \exp\left(\frac{U(a^*(s),\theta)}{\lambda}\right) ds. \tag{10}$$

This is the Gibbs distribution of statistical mechanics, with the information price λ acting as an inverse temperature: low λ corresponds to precise, sharply discriminating signals, while

[5] Berk (1997) establishes an analogous necessity result in the CAPM setting: quadratic utility is necessary, not merely sufficient, for mean-variance preferences under expected utility maximisation, just as each of our three conditions (G1)–(G3) is necessary for mean-variance sufficiency of the VoI.

[6] (Stratonovich, 1975/2020, Section 3.4) writes $Z(\beta)$ suppressing the state argument as works at the level of the *expected* free energy.

high λ yields diffuse signals. Comparing with Definition (4), the free energy $F(\theta) = \lambda \log Z(\theta)$ of state θ is precisely the certainty equivalent of the utility lottery over signals in that state,

$$F(\theta) = \lambda \log Z(\theta) = \lambda \log \int p(s) \exp\left(\frac{U(a^*(s), \theta)}{\lambda}\right) ds = \mathrm{CE}_\lambda\big(U(a^*(\cdot), \theta)\big), \tag{11}$$

where the expectation inside CE_λ is taken over signals s with marginal distribution $p(s)$.

5.2. *Free Energy and the Primal-Dual Identity*

LEMMA 2—Primal-Dual Identity: *At the interior optimum with information price $\lambda^*(\bar{I}) = V'(\bar{I})$,*

$$U(\bar{I}) \;=\; E_\theta\big\{F(\theta)\big\} + \lambda^*\, \bar{I}, \tag{12}$$

where $F(\theta) = \mathrm{CE}_{\lambda^*}(U(a^*(\cdot), \theta))$ is the free energy and λ^* is chosen so that $I(\Theta, S) = \bar{I}$ under the Gibbs signal.

The value function is the prior expectation of the state-level certainty equivalent plus the total shadow cost $\lambda^* \bar{I}$ of information capacity. Combining with (3):

$$V(\bar{I}) = E_\theta\big\{F(\theta)\big\} + \lambda^* \bar{I} - U(0), \tag{13}$$

where $U(0) = E_\theta\{U(a_0^*, \theta)\}$ and $a_0^* = \arg\max_a E\{U(a, \Theta)\}$ is the prior-optimal action with no signal. This generalises the Gaussian quadratic formula of Lemma 1: in the Gaussian quadratic case the free energy reduces to an affine function of posterior variance and (13) collapses to (7).

The partition function $Z(\theta)$ plays a decisive role in Theorem 4: it is state-independent under Gaussian quadratic structure (condition G2), which eliminates state-dependent tail weighting and makes the VoI depend only on posterior variance. Outside the knife-edge, $Z(\theta)$ varies across states, generating asymmetric concentration of capacity on states where the optimal action differs most from the exit option, as the portfolio application demonstrates.

6. THE VOI RISK MEASURE

Since the VoI $V(\bar{I})$ is strictly concave in $\bar{I}$, its Legendre–Fenchel transform is well-defined and strictly convex in λ. The information price $\lambda^*(\bar{I}) = V'(\bar{I})$ follows from the envelope theorem, establishing duality between information levels and information prices as characterised by Theorem 3.

DEFINITION 3—VoI Risk Measure: *The Stratonovich/Shannon VoI risk measure is the convex conjugate of the value function:*

$$\mathcal{R}_\lambda^{\mathrm{VoI}}(X) = \sup_{I \geq 0} \big\{V(I; X) - \lambda\, I\big\}, \tag{14}$$

where $V(I; X)$ is the value function for random variable X under mutual information constraint I, and λ is the information price. Equivalently, by Theorem 1:

$$\mathcal{R}_\lambda^{\mathrm{VoI}}(X) = \max_{p(s|\theta)} \big\{E\{U(a^*(S), \Theta)\} - \lambda\, I(\Theta, S)\big\} = V_{\mathrm{soft}}(\lambda).$$

Geometrically, $\mathcal{R}_\lambda^{\mathrm{VoI}}$ equals the vertical intercept of the tangent line to $V(I)$ with slope λ. Since $V(I)$ is strictly concave, the risk measure is strictly convex in λ: higher information prices generate higher measured risk.

THEOREM 5—Properties of the VoI Risk Measure: *The VoI risk measure $\mathcal{R}_\lambda^{\mathrm{VoI}}$ satisfies:*

(i) Expected entropic representation: $\mathcal{R}_\lambda^{\mathrm{VoI}}(X) = E_S\big\{\mathcal{R}_\lambda(U(\cdot,S))\big\}$, *where* $\mathcal{R}_\lambda(Y) = \lambda \log E\{\exp(Y/\lambda)\}$ *is the entropic risk measure.*

(ii) Extended coherence (Rockafellar, 2007): *monotonicity, convexity, and closedness hold.*

(iii) Failure of positive homogeneity: $\mathcal{R}_\lambda^{\mathrm{VoI}}(\alpha X) \neq \alpha\, \mathcal{R}_\lambda^{\mathrm{VoI}}(X)$ *for* $\alpha > 0$, $\alpha \neq 1$.

Proofs of Theorem 5 and the full axiomatic development of the VoI risk measure class appear in Behringer (2026). The expected entropic representation reveals that the VoI framework generalises entropic risk: instead of a single exponential moment, we take expectations over signal realisations. The failure of positive homogeneity in (iii) is a structural consequence of the non-homogeneity of Shannon mutual information. It generates empirically testable scale effects: scaling up exposure to a risky prospect increases optimal information expenditure more than proportionately. The properties in (ii) coincide with the definition of a *subregular* risk measure in the sense of Grechuk, Malandii, Rockafellar, and Uryasev (2026), who develop the axiomatic and duality theory for this class.

By Legendre–Fenchel duality and Lemma 2, $V(\bar{I}; X) = \inf_\lambda\{\mathcal{R}_\lambda^{\mathrm{VoI}}(X) + \lambda \bar{I}\}$. Hence if $V(\bar{I}; X) > V(\bar{I}; Y)$ for all $\bar{I} \geq 0$, then $\mathcal{R}_\lambda^{\mathrm{VoI}}(X) > \mathcal{R}_\lambda^{\mathrm{VoI}}(Y)$ for all $\lambda > 0$: the VoI risk measure assigns uniformly higher risk to any random variable whose VoI exceeds that of another at every information budget. This linkage is exploited in Application 2 below.

7. APPLICATIONS

7.1. *Medical Diagnostic Testing*

A physician faces a binary diagnostic problem (Stratonovich, 1975/2020, Section 10.1): a patient either has a disease (θ_1) or does not (θ_2). The physician chooses between treating (a_1) or not treating (a_2). A diagnostic test with sensitivity $\mathrm{se} = P(S = 1|\theta_1)$ and specificity $\mathrm{sp} = P(S = 0|\theta_2)$ produces a signal S at cost $c > 0$ per patient. Under the identity utility matrix as a symmetric benchmark (for arbitrary matrices see Behringer and Belavkin (2026)), the VoI is independent of disease prevalence. Evaluating at equal priors $p = \frac{1}{2}$:

$$I(D,S) = H\left(\tfrac{\mathrm{se}+1-\mathrm{sp}}{2}\right) - \tfrac{1}{2}H(\mathrm{se}) - \tfrac{1}{2}H(1-\mathrm{sp}), \tag{15}$$

where $H(x) = -x\ln x - (1-x)\ln(1-x)$ is binary entropy.

The physician orders the test if and only if $V(I(D,S)) \geq c$ where V is the VoI function defined in (3). Since V is strictly increasing in $I(D,S)$, there exists a unique threshold I^* satisfying $V(I^*) = c$, with the test ordered whenever $I(D,S) \geq I^*$. Uniqueness follows from strict concavity of V and the intermediate value theorem.

The framework also determines willingness to pay for additional diagnostic precision: if an improved test raises informativeness from I_0 to I_1, the maximum upgrade value is $V(I_1) - V(I_0)$. By strict concavity this is strictly decreasing in I_0: a given precision gain $I_1 - I_0$ is worth strictly more when baseline accuracy is low than when it is high, providing a direct decision rule for R&D investment in diagnostics. This result extends to asymmetric utility matrices capturing different costs of false positives and false negatives, with the threshold depending on the prior, the full matrix, se, sp, and c.

7.2. *Portfolio Choice Under Tail Risk*

An investor with CARA utility $U(w) = -e^{-\gamma w}$ ($\gamma = 2$) allocates between a riskless asset ($r_f = 2\%$) and a risky asset with excess return $\tilde{R} - r_f$. We construct two discrete distributions with *identical* means (8%) and standard deviations (18%) but different higher moments, calibrated to S&P 500 return data (1950–2024, mean 8%, volatility 18%, skewness -0.5, excess kurtosis 3):

TABLE I

SYMMETRIC AND SKEWED RETURN DISTRIBUTIONS WITH IDENTICAL MEANS AND STANDARD DEVIATIONS

Distribution	States	Returns	Probabilities
Symmetric	Loss, Gain	-10%, $+26\%$	50%, 50%
Skewed	Crash, Normal, Gain	-40%, $+8.8\%$, $+30\%$	10%, 70.75%, 19.25%

The skewed distribution has skewness -1.54 and excess kurtosis 2.49; both means and variances are matched exactly. The binary action space ($\alpha \in \{0,1\}$, full investment or full allocation to the riskless asset) keeps the signal design problem explicit while capturing the essential portfolio decision of whether to be in or out of equities. All quantities follow analytically from the Gibbs fixed-point system (Proposition 3) and the primal-dual identity (Lemma 2); detailed derivations are in the Online Appendix.

TABLE II

STRATONOVICH/SHANNON VOI AT FIXED INFORMATION LEVELS ($\gamma = 2$, $r_f = 2\%$, $\alpha \in \{0,1\}$)

$\bar{I}$ (bits)	VoI Sym (%)	VoI Skew (%)	λ^*_{Sym}	λ^*_{Skew}	Ratio
0.10	2.64	5.90	0.303	0.451	2.24
0.20	4.45	8.92	0.229	0.346	2.00
0.30	5.88	10.86	0.186	0.218	1.85

At $\bar{I} = 0.30$ bits, the VoI is 5.88% of wealth under the symmetric distribution and 10.86% under the skewed distribution, a ratio of 1.85 or 85% higher. The mechanism is transparent from the Gibbs optimal signal (9): the crash state (-40%) concentrates 81.5% of the signal's mutual information capacity despite having only 10% prior probability (Table III).

The key mechanism is the exponent $(u_0 - u_1^{\text{crash}})/\lambda^* = 1.18/0.218 \approx 5.4$, giving $e^{5.4} \approx 221$: the Gibbs distribution places 221 times more weight on the exit action than on the stay action in the crash state, reflecting the large utility gap between investing in a crash and holding the riskless asset. Combined with the unconditional exit probability $q^* \approx 0.163$, this produces an exit signal probability of 0.982, i.e. near certainty in the crash state. This concentration is driven by the utility gap between the invest and exit actions in the crash state, not by the prior probability.

Implications for mean-variance analysis. Theorem 4 establishes that information value depends only on posterior variance if and only if all three Gaussian quadratic conditions hold. Both distributions have identical means and variances, so a Gaussian quadratic model assigns

TABLE III

OPTIMAL SIGNAL CAPACITY ALLOCATION BY STATE AT $\bar{I} = 0.30$ BITS

State	Symmetric	Skewed
Crash / Loss	41.6%	81.5%
Normal	—	8.9%
Gain	58.4%	9.6%
P(exit signal | downside state)	63.4%	98.2%

identical information value to both via (7). In reality, the skewed distribution makes crash-detecting information 85–124% more valuable depending on the information budget. Standard finance models that evaluate data acquisition using symmetric benchmarks, whether through explicit Gaussian assumptions or implicit reliance on variance as a sufficient statistic, therefore systematically underprice tail-sensitive information. The bias is largest when information budgets are tightest, precisely the regime where optimal allocation matters most.

By Theorem 5, since $V(\bar{I}; X_{\text{skewed}}) > V(\bar{I}; X_{\text{symmetric}})$ for all $\bar{I} \geq 0$, the VoI risk measure assigns uniformly higher risk to the negatively skewed distribution: when crash-detecting capacity is limited, downside states are more costly under skewness, a distinction that variance-based risk measures cannot capture. The failure of positive homogeneity implies that as the exposure to the skewed distribution scales up, the VoI risk measure grows more than proportionately, reflecting the exponential Gibbs weighting which assigns progressively larger weight to downside states.

Nonlinear scaling with risk aversion. Table IV documents that when γ doubles, VoI more than doubles, with the effect stronger for the skewed distribution.

TABLE IV

NONLINEAR SCALING OF VOI WITH RISK AVERSION AT $\bar{I} = 0.30$ BITS

γ	VoI Symmetric (%)	VoI Skewed (%)	Ratio
1	2.71	4.22	1.56
2	5.88	10.86	1.85
4	14.12	31.86	2.26

This nonlinear scaling is the failure of positive homogeneity (Theorem 5(iii)) made quantitative: higher risk aversion assigns exponentially greater Gibbs weight to crash states, disproportionately increasing optimal information expenditure.

7.3. *Climate Policy Under Deep Uncertainty*

Climate policy confronts a fundamental challenge: key parameters such as climate sensitivity and damage functions are subject to profound expert disagreement (Heal and Millner, 2014). The VoI risk measure handles risk and uncertainty simultaneously within a single framework.

A policymaker chooses mitigation level a given uncertain climate sensitivity Θ and damages $U(a, \Theta)$. By Theorem 5(i),

$$\mathcal{R}_\lambda^{\text{VoI}}(U) = E_S\{\mathcal{R}_\lambda(U(\cdot, S))\},$$

where $\mathcal{R}_\lambda(Y) = \lambda \log E\{\exp(Y/\lambda)\}$. This representation separates the two sources of uncertainty cleanly: probabilistic risk over damage outcomes given a signal realisation enters through $\mathcal{R}_\lambda(U(\cdot, S))$, while uncertainty about the information structure enters through the outer expectation over S. The information price λ mediates between these continuously: as $\lambda \to \infty$ the policymaker approaches risk neutrality; as $\lambda \to 0$ the policymaker places increasing weight on worst-case outcomes, approaching behaviour associated with deep uncertainty without requiring a separate maxmin operator or a set of priors.

The hard information constraint $I(\Theta, S) \leq \bar{I}$ applies naturally in this context: climate research capacity faces genuine institutional and technological limits rather than smoothly adjustable marginal costs, making the hard formulation the economically appropriate one. The shadow price $\lambda^*(\bar{I}) = V'(\bar{I})$ equals the marginal value of expanding research capacity, providing a decision-theoretic foundation for climate science funding: by the primal-dual identity (Lemma 2), the value of an incremental expansion of the research budget by $d\bar{I}$ is exactly $\lambda^* \, d\bar{I}$, denominated in the same units as climate damages.

The non-homogeneity of $\mathcal{R}_\lambda^{\mathrm{VoI}}$ has a direct interpretation: scaling up the severity of climate damages disproportionately increases the VoI, reflecting a superlinear interaction between tail risk and information acquisition that is absent from homogeneous risk measures (Pindyck, 2013).

8. CONCLUSION

The structural advantage of the hard-constraint formulation of rational inattention has a precise counterpart in convex duality theory. The soft-constraint objective is the Fenchel conjugate of the VoI, fully recoverable by the biconjugate theorem (Theorem 2). The marginal cost is identified without parametric restriction by the inverse subgradient theorem (Theorem 3). The linear entropy penalty is a special case, not a maintained assumption, and is inappropriate where information costs are decision-problem specific and non-linear in bits (Sims, 2006).

The Gaussian quadratic benchmark, the workhorse of the applied RI literature (see, e.g., (Maćkowiak, Matějka, and Wiederholt, 2023)), is characterised as a knife-edge by Theorem 4: mean-variance sufficiency of the VoI requires conjugacy, state independence, and second-moment sufficiency to hold simultaneously, and each condition is necessary. In the portfolio application calibrated to S&P 500 return moments, removing the symmetry assumption while keeping means and variances identical raises the VoI by 85–124% and concentrates 81.5% of optimal signal capacity on crash detection, a pattern that standard mean-variance analysis cannot produce.

The VoI risk measure (Definition 3) connects the value function to risk measure theory via the primal-dual identity (Lemma 2). It satisfies extended coherence (Theorem 5) and provides a tractable framework for optimal information acquisition when capacity is a hard technological or institutional limit. The Lagrangian dual serves as a computational tool to characterise optimal signals, but the underlying economic problem remains one of constrained capacity rather than marginally priced information. The rate distortion approach of Stratonovich (1975/2020), combined with the convex duality results of Rockafellar (1974), provides the natural mathematical foundation for rational inattention models with non-linear information costs.

Two extensions are natural. The continuous-state case with non-Gaussian priors would sharpen the quantitative predictions beyond the discrete calibration developed here. Multi-period settings would connect the hard-constraint formulation to dynamic RI models, with the shadow price evolving as capacity constraints bind sequentially, and the VoI risk measure to dynamic risk measures.

APPENDIX A: PROOFS

PROOF OF THEOREM 1: We establish both inequalities.

Step 1 (weak duality, $\geq$). Fix $\lambda \geq 0$ and $\bar{I} \geq 0$. For any signal structure $p(s|\theta)$ satisfying $I(\Theta, S) \leq \bar{I}$,

$$E\{U\} - \lambda I(\Theta, S) + \lambda \bar{I} \;\geq\; E\{U\}.$$

Taking the supremum over all such signal structures gives $V_{\text{soft}}(\lambda) + \lambda \bar{I} \geq V(\bar{I})$, hence

$$V_{\text{soft}}(\lambda) \;\geq\; \sup_{\bar{I} \geq 0} \{V(\bar{I}) - \lambda \bar{I}\}.$$

Step 2 (reverse inequality, $\leq$). Let $\{p_n\}$ be a maximising sequence for the soft problem, with $I_n = I(\Theta, S)$ under p_n. Feasibility of p_n for the hard problem at budget I_n gives $E\{U\}_n \leq V(I_n)$, so

$$E\{U\}_n - \lambda I_n \;\leq\; V(I_n) - \lambda I_n \;\leq\; \sup_{\bar{I} \geq 0} \{V(\bar{I}) - \lambda \bar{I}\}.$$

Taking limits: $V_{\text{soft}}(\lambda) \leq \sup_{\bar{I} \geq 0} \{V(\bar{I}) - \lambda \bar{I}\}$.

Both steps together establish the result. *Q.E.D.*

PROOF OF THEOREM 2: By Theorem 1, $V_{\text{soft}}(\lambda) = \sup_{\bar{I}} \{V(\bar{I}) - \lambda \bar{I}\}$, which is the Fenchel conjugate of $-V(\bar{I})$ evaluated at $-\lambda$ (or equivalently the concave conjugate of V at λ). Since $V(\bar{I})$ is continuous and strictly concave under Assumption 1, $-V(\bar{I})$ is closed and convex. By Rockafellar (1974, Theorem 5, p. 16), $\operatorname{cl}\operatorname{co}(-V) = -V$, so the biconjugate satisfies $(-V)^{**} = -V$. Unfolding the biconjugate in terms of V_{soft}:

$$(-V)^{**}(\bar{I}) = \sup_{\lambda} \big\{\lambda \bar{I} - (-V)^*(\lambda)\big\} = \sup_{\lambda} \big\{\lambda \bar{I} - V_{\text{soft}}(\lambda) + \text{const}\big\},$$

or equivalently $V(\bar{I}) = \inf_{\lambda \geq 0} \{V_{\text{soft}}(\lambda) + \lambda \bar{I}\}$, where the infimum is attained at $\lambda^* = V'(\bar{I})$ by Theorem 3. The recovery requires no parametric assumption: given $V_{\text{soft}}(\lambda)$ as a function of λ, $V(\bar{I})$ is recovered pointwise by this infimum for any $\bar{I}$. *Q.E.D.*

PROOF OF THEOREM 3: Under Assumption 1, $-V(\bar{I})$ is a closed convex function with non-empty subdifferential at every interior $\bar{I}$. By Rockafellar (1974, Theorem 12, p. 35) applied to $\varphi = -V$, $u = \bar{I}$, $y = -\lambda^*$:

$$\partial(-V)(\bar{I}) = \big\{y \mid \bar{I} \in \partial(-V)^*(y)\big\}.$$

Since $(-V)^*(\lambda) = V_{\text{soft}}(\lambda)$ by Theorem 1 (up to the sign convention for the conjugate), translating:

$$-\partial V(\bar{I}) = \big\{-\lambda^* \mid \bar{I} \in -\partial V_{\text{soft}}(\lambda^*)\big\},$$

which gives $\partial V(\bar{I}) = \{\lambda^* \mid \bar{I} \in \partial V_{\text{soft}}(\lambda^*)\}$. When V is differentiable at $\bar{I}$, the subdifferential is a singleton, $\partial V(\bar{I}) = \{V'(\bar{I})\}$, so $\lambda^*(\bar{I}) = V'(\bar{I})$. Symmetrically, when V_{soft} is differentiable at λ^*, $\partial V_{\text{soft}}(\lambda^*) = \{V'_{\text{soft}}(\lambda^*)\} = \{-\bar{I}^*\}$ by Theorem 1, establishing the mutual identification of primal and dual solutions. *Q.E.D.*

PROOF OF LEMMA 1: Under no signal, the optimal action is the prior mean and $U_0 = -\sigma_\theta^2$. With signal S achieving mutual information I, the posterior variance is $\text{Var}(\Theta|S) =$

$\sigma_\theta^2 e^{-2I}$ and expected utility equals minus the posterior variance, $U_I = -\sigma_\theta^2 e^{-2I}$. Hence $V(I) = U_I - U_0 = -\sigma_\theta^2 e^{-2I} - (-\sigma_\theta^2) = \sigma_\theta^2(1 - e^{-2I})$. Strict concavity follows from $V''(I) = -4\sigma_\theta^2 e^{-2I} < 0$. *Q.E.D.*

PROOF OF THEOREM 4: **(ii) ⇒ (i).** Let $U(a,\theta) = -(a-\theta)^2$ and $\theta \sim \mathcal{N}(\mu_0, \sigma_0^2)$. With Gaussian signals, Bayes' rule implies that posteriors remain Gaussian, establishing (G1). Under quadratic loss, the optimal action equals the posterior mean, and conditional expected utility equals minus the posterior variance, so expected utility depends only on first and second moments, establishing (G3). Because the exponent in the optimal signal distribution is quadratic in θ, integration yields a normalising constant independent of θ, establishing (G2).

(i) ⇒ (iii). Under Gaussian quadratic structure, posterior beliefs are fully characterised by mean and variance (G1), utility depends only on these moments (G3), and the signal normalisation does not distort tail weights across states (G2). Hence information affects expected utility only through posterior variance reduction, so the VoI depends solely on variance reduction.

(iii) ⇒ (ii). Suppose the VoI depends only on posterior variance uniformly over all budgets $\bar{I} \in (0, C)$ and all feasible signal distributions.

Step 1 (Utility). If expected utility under the optimal action depends only on posterior variance for all feasible posteriors, then the map $\sigma^2 \mapsto \max_a E\{U(a,\Theta) \mid \mathrm{Var}(\Theta) = \sigma^2\}$ must be well-defined. Under Assumption 1(R3), the first-order condition for the optimal action pins it to the posterior mean, and the envelope theorem implies that the variation of expected utility with respect to posterior moments beyond the second must vanish for all feasible posteriors. This forces the utility function to be at most quadratic in θ; positive affine transformations do not affect optimality, giving $U(a,\theta) = -(a-\theta)^2$ up to such a transformation.

Step 2 (Prior and signals). If the VoI depends only on posterior variance for every budget $\bar{I} \in (0, C)$, then under the Shannon mutual information constraint the posterior family must be closed and fully characterised by its first two moments, which within continuous distributions satisfying Assumption 1(R1)–(R2) is consistent only with the Gaussian family (Cover and Thomas, 2006, Theorem 8.6.5), establishing (G1). State independence of the normalising constant (G2) then follows because under quadratic utility and Gaussian signals the exponent in the Gibbs distribution is quadratic in θ, so integration over signals yields a state-independent constant.

Final claim. Failure of (G1) introduces higher-order posterior moments that are not suppressed by the Shannon mutual information constraint; failure of (G2) creates state-dependent tail weighting through the state-dependent normalising constant; and failure of (G3) makes utility directly sensitive to moments beyond variance. In each case, for generic parameter values, the VoI depends on posterior moments beyond variance. Non-generic exceptions occur only when different sources of non-Gaussianity cancel in the normalising constant, a set of measure zero in the space of priors and utility functions. *Q.E.D.*

PROOF OF PROPOSITION 3: Taking the variational derivative of the Lagrangian $\mathcal{L}[p(s|\theta)] = E\{U(a^*(S),\Theta)\} - \lambda I(\Theta, S)$ with respect to $p(s|\theta)$ and setting it to zero, subject to the normalisation constraint $\int p(s|\theta)\, ds = 1$, yields the first-order condition

$$U(a^*(s),\theta) - \lambda \log \frac{p(s|\theta)}{p(s)} - \lambda - \mu(\theta) = 0,$$

where $\mu(\theta)$ is the Lagrange multiplier on normalisation. Solving for $p(s|\theta)$ gives (9) with $Z(\theta) = e^{1+\mu(\theta)/\lambda}$ identified by normalisation as in (10). *Q.E.D.*

PROOF OF LEMMA 2: At the Gibbs optimum $p^*(s|\theta) = p(s)e^{U/\lambda}/Z(\theta)$, substituting into $E\{U\} - \lambda I$ gives:

$$\begin{aligned} E\{U\} - \lambda I &= \sum_\theta \mu(\theta) \sum_s p^*(s|\theta) \Big[U(a^*(s),\theta) - \lambda \log \tfrac{p^*(s|\theta)}{p(s)} \Big] \\ &= \sum_\theta \mu(\theta) \sum_s p^*(s|\theta) \Big[U(a^*(s),\theta) - U(a^*(s),\theta) + \lambda \log Z(\theta) \Big] \\ &= \sum_\theta \mu(\theta)\, \lambda \log Z(\theta) = E_\theta\big\{F(\theta)\big\}, \end{aligned}$$

where $F(\theta) = \lambda \log Z(\theta) = \mathrm{CE}_\lambda(U(a^*(\cdot),\theta))$ is the free energy. By Theorem 1 and Theorem 2, $U(\bar{I}) = \inf_{\lambda \geq 0}\{V_{\text{soft}}(\lambda) + \lambda \bar{I}\} + U(0)$, and at the interior optimum where the infimum is attained at $\lambda^* = \bar{V}'(\bar{I})$, evaluating $V_{\text{soft}}(\lambda^*) = E_\theta\{F(\theta)\}|_{\lambda=\lambda^*}$ from the expression above and using $U(\bar{I}) = V_{\text{soft}}(\lambda^*) + \lambda^* \bar{I}$ gives (12). For example in (Stratonovich, 1975/2020, Section 9.4). *Q.E.D.*

APPENDIX: REFERENCES

Azrieli, Yaron. 2021. "Constrained versus Unconstrained Rational Inattention." *Games* 12 (3): 1–22. [1, 5]

Behringer, Stefan. 2026. "Value of Information Based Risk Measures: Theory and Applications." mimeo. [10]

Behringer, Stefan, and Roman V. Belavkin. 2023. "The Value of Information and Circular Settings." arXiv:2303.16126 [econ.TH]. [2]

Behringer, Stefan, and Roman V. Belavkin. 2025. "The Value of Information in Economic Contexts." *Physical Sciences Forum* 12 (1): 6. [2]

Behringer, Stefan, and Roman V. Belavkin. 2026. "Value of Information in Bayesian Environments." In *Dynamics of Information Systems*, edited by Hossein Moosaei, Roman Belavkin, and Panos M. Pardalos, 16–24. Springer Nature Switzerland. [10]

Berk, Jonathan B. 1997. "Necessary Conditions for the CAPM." *Journal of Economic Theory* 73 (1): 245–57. [8]

Bloedel, Alexander W., Tommaso Denti, and Luca Pomatto. 2025. "Modeling Information Acquisition via f-Divergence and Duality.", arXiv:2510:03482 [econ.TH]. [3]

Cover, Thomas M., and Joy A. Thomas. 2006. *Elements of Information Theory*. 2nd ed. Hoboken: Wiley. [7, 15]

Denti, Tommaso, Massimo Marinacci, and Luigi Montrucchio. 2020. "A Note on Rational Inattention and Rate Distortion Theory." *Decisions in Economics and Finance* 43: 75–89. [2]

Denti, Tommaso, Massimo Marinacci, and Aldo Rustichini. 2022. "Experimental Cost of Information." *American Economic Review* 112 (9): 3106–23. [2]

Föllmer, Hans, and Alexander Schied. 2004. *Stochastic Finance: An Introduction in Discrete Time*. 2nd ed. Berlin: de Gruyter. [2]

Fosgerau, Mogens, Emerson Melo, André de Palma, and Matthew Shum. 2020. "Discrete Choice and Rational Inattention: A General Equivalence Result." *International Economic Review* 61 (4): 1569–89. [3]

Gershkov, Alex, Benny Moldovanu, Philipp Strack, and Mengxi Zhang. 2025. "Optimal Security Design for Risk-Averse Investors." *American Economic Review* 115 (6): 2050–92. [2]

Grechuk, Bogdan, Anton Malandii, R. Tyrrell Rockafellar, and Stanislav Uryasev. 2026. "The Risk Quadrangle in Optimization: An Overview with Recent Results and Extensions." mimeo. [10]

Hansen, Lars Peter, and Thomas J. Sargent. 2001. "Robust Control and Model Uncertainty." *American Economic Review* 91 (2): 60–66. [2]

Hansen, Lars Peter, and Thomas J. Sargent. 2008. *Robustness*. Princeton: Princeton University Press. [2]

Heal, Geoffrey, and Antony Millner. 2014. "Uncertainty and Decision Making in Climate Change Economics." *Review of Environmental Economics and Policy* 8 (1): 120–37. [12]

De Lara, Michel, and Olivier Gossner. 2020. "Payoffs-Beliefs Duality and the Value of Information." *SIAM Journal on Optimization* 30 (1): 464–489. [2]

Maćkowiak, Bartosz, Filip Matějka, and Mirko Wiederholt. 2023. "Rational Inattention: A Review." *Journal of Economic Literature* 61 (1): 226–73. [1, 13]

Matějka, Filip, and Alisdair McKay. 2015. "Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model." *American Economic Review* 105 (1): 272–98. [2]

De Oliveira, Henrique, Tommaso Denti, Maximilian Mihm, and Kemal Ozbek. 2017. "Rationally Inattentive Preferences and Hidden Information Costs." *Theoretical Economics* 12: 621–54. [3]

Pindyck, Robert S. 2013. "Climate Change Policy: What Do the Models Tell Us?" *Journal of Economic Literature* 51 (3): 860–72. [13]

Rockafellar, R. Tyrrell. 1974. *Conjugate Duality and Optimization*. CBMS-NSF Regional Conference Series in Applied Mathematics, Vol. 16. Philadelphia: Society for Industrial and Applied Mathematics. [6, 13, 14]

Rockafellar, R. Tyrrell. 2007. "Coherent Approaches to Risk in Optimization Under Uncertainty." In *Tutorials in Operations Research, INFORMS*, 38–61. [2, 10]

Shannon, Claude E. 1948. "A Mathematical Theory of Communication." *Bell System Technical Journal* 27 (3): 379–423. [7]

Sims, Christopher A. 2003. "Implications of Rational Inattention." *Journal of Monetary Economics* 50 (3): 665–90. [1, 2, 6, 7]

Sims, Christopher A. 2006. "Rational Inattention: Beyond the Linear-Quadratic Case." *American Economic Review* 96 (2): 158–63. [2, 6, 7, 13]

Stratonovich, Ruslan L. 1975/2020. *Theory of Information and Its Value*, edited by Roman V. Belavkin, Panos M. Pardalos, and Jose C. Principe. Cham: Springer. [1, 2, 4, 8, 10, 13, 16]

Van Nieuwerburgh, Stijn, and Laura Veldkamp. 2010. "Information Acquisition and Under-Diversification." *Review of Economic Studies* 77 (2): 779–805. [2]