

VALUE OF INFORMATION BASED RISK MEASURES: THEORY AND APPLICATIONS

STEFAN BEHRINGER

Value of Information Group, Universität Bielefeld

The value of a decision problem to a rationally inattentive agent is a functional of its payoffs. Under a linear penalty on Shannon mutual information this functional is the maximum over action marginals of an expected entropic certainty equivalent; its reflection is a convex risk measure, which fails positive homogeneity exactly when information has value. Under a hard bound on mutual information the functional is positively homogeneous, its reflection a coherent risk measure of the entropic value-at-risk type, and the conjugate duality between the two formulations is the dual formula of the entropic value-at-risk. The information premium, the excess of the penalised value over the value without information, is invariant to translations, falls with the price of information, grows more than proportionately with the stakes and orders decision problems by their value of information. Applications to diagnostic testing, portfolio choice under tail risk and climate research show how a binding capacity concentrates information on states with large payoff gaps rather than large prior probability.

JEL CLASSIFICATION. D81, D83, G11, G32.

KEYWORDS: Value of information, rational inattention, entropic risk measure, entropic value-at-risk, convex duality.

1. INTRODUCTION

Rational inattention (RI) has two formulations: a hard bound on the Shannon mutual information that an agent can process (Sims, 2003), and a linear penalty on it (Matějka and McKay, 2015). Behringer (2026) studies the two through the Stratonovich/Shannon Value of Information (VoI), the gain in expected utility attainable with a given capacity (Stratonovich, 1975/2020), and shows that the penalised value is the Fenchel conjugate of the VoI. This paper reads both formulations as functionals of the payoffs of a decision problem and relates them to the risk measures of mathematical finance.

The connection rests on known results. The penalised value is an expected logsum (Matějka and McKay, 2015, Denti, Marinacci, and Montrucchio, 2020), which is an entropic certainty equivalent, and the variational formula behind it (Donsker and Varadhan, 1975) is the dual representation of the entropic risk measure (Föllmer and Schied, 2016). What the paper adds is a classification of the two RI values as risk functionals (Proposition 2): the penalised value is convex, monotone and cash-additive but not positively homogeneous, the constrained value is in addition positively homogeneous, and the conjugate duality between them is the dual formula of the entropic value-at-risk of Ahmadi-Javid (2012). The difference between the formulations in their response to the stakes, studied in Behringer (2026, Section 6), is the difference between a convex and a coherent risk measure.

The paper also studies the information premium, the excess of the penalised value over the value without information (Proposition 3). It is not itself a risk measure, but it orders decision problems by their VoI, grows more than proportionately with the stakes and falls with the price of information. Three applications illustrate the framework: diagnostic testing, portfolio choice under tail risk, and the allocation of research capacity in climate policy. In each, a binding capacity directs information to the states with the largest payoff gaps, which need not be the most probable.

Section 2 sets up the decision problem. Section 3 gives the entropic representation, Section 4 the classification and the information premium, and Section 5 the applications. Section 6 concludes. Appendix A collects the proofs and Appendix B derives the numbers of the portfolio application.

2. SETUP

There is a finite set Ω of states with prior μ , assumed to have full support, and a finite set $\mathcal{A}$ of actions.

DEFINITION 1—Decision problem, values and VoI: A *decision problem* is a payoff matrix $u \in \mathbb{R}^{\mathcal{A} \times \Omega}$, with $u(a, \theta)$ the utility of action a in state θ . A *channel* q assigns to each state θ a distribution $q(\cdot \mid \theta)$ on $\mathcal{A}$, with action marginal $P_q(a) := \sum_{\theta} \mu(\theta) q(a \mid \theta)$, expected

payoff

$$\langle q, u \rangle := \sum_{\theta} \mu(\theta) \sum_a q(a | \theta) u(a, \theta)$$

and mutual information $I(q) := \sum_{\theta} \mu(\theta) \sum_a q(a | \theta) \ln(q(a | \theta) / P_q(a))$. For $\lambda > 0$ and $\bar{I} \geq 0$ the *penalised* and *constrained* values are

$$\Phi_{\lambda}(u) := \max_q \{ \langle q, u \rangle - \lambda I(q) \}, \quad \Psi_{\bar{I}}(u) := \max_q \{ \langle q, u \rangle : I(q) \leq \bar{I} \}.$$

The *value without information* is $U_u(0) := \Psi_0(u) = \max_a \sum_{\theta} \mu(\theta) u(a, \theta)$, and the *VoI* is $V_u(\bar{I}) := \Psi_{\bar{I}}(u) - U_u(0)$.

A channel recommends actions, which is without loss of generality (Behringer, 2026, Lemma 3). The VoI is non-decreasing, concave and continuous in $\bar{I}$, and $\Phi_{\lambda}(u) = U_u(0) + \sup_{\bar{I} \geq 0} \{ V_u(\bar{I}) - \lambda \bar{I} \}$ (Behringer, 2026, Theorem 1 and Lemma 4).

3. THE ENTROPIC REPRESENTATION

PROPOSITION 1—Entropic representation: *For every $\lambda > 0$,*

$$\Phi_{\lambda}(u) = \max_{P \in \Delta(\mathcal{A})} \sum_{\theta} \mu(\theta) W_{\lambda}(\theta; P), \quad W_{\lambda}(\theta; P) := \lambda \ln \sum_a P(a) e^{u(a, \theta) / \lambda},$$

and for a maximising P the optimal channel is the Gibbs kernel

$$q(a | \theta) = \frac{P(a) e^{u(a, \theta) / \lambda}}{\sum_b P(b) e^{u(b, \theta) / \lambda}}, \quad (1)$$

whose action marginal is P .

For Shannon costs the representation is due to Denti, Marinacci, and Montrucchio (2020, Theorem 1 and Lemma 3), and the Gibbs form of the solution to Matějka and McKay (2015); it is the expected logsum of Behringer (2026, Section 5). The inner expectation in $W_{\lambda}(\theta; P)$ is over actions drawn from the *unconditional* action distribution P , not over the signals received in state θ ; the outer expectation is over states.

REMARK—Risk attitude: $W_{\lambda}(\theta; P)$ is the certainty equivalent of the payoff $u(A, \theta)$ of an action $A \sim P$, evaluated with the convex utility $e^{z/\lambda}$, and so exceeds its ex-

pectation by Jensen's inequality. The informed agent values the dispersion of payoffs across actions, because information lets her select among them: the excess of W_λ over $\sum_a P(a)u(a, \theta)$ is the option value of information. It vanishes as $\lambda \rightarrow \infty$ and approaches $\max_{a: P(a) > 0} u(a, \theta) - \sum_a P(a)u(a, \theta)$ as $\lambda \rightarrow 0$. Reflecting the payoff gives the familiar risk-averse objects: W_λ evaluated at $-u$ is the entropic risk measure of u with risk tolerance λ (Föllmer and Schied, 2016), and $-W_\lambda$ evaluated at $-u$ is the multiplier criterion of robust control (Hansen and Sargent, 2001), in which the distortion of beliefs is chosen to minimise rather than to maximise the expected payoff.

4. RISK MEASURES

4.1. Classification

A functional ϕ on $\mathbb{R}^{A \times \Omega}$ is *monotone* if $u \leq u'$ implies $\phi(u) \leq \phi(u')$, *cash-additive* if $\phi(u + c) = \phi(u) + c$ for every constant c , and *positively homogeneous* if $\phi(ku) = k\phi(u)$ for every $k > 0$. If ϕ is monotone, convex and cash-additive with $\phi(0) = 0$, then $\rho(x) := \phi(-x)$ is a convex risk measure on positions x in the sense of Föllmer and Schied (2002) and Frittelli and Rosazza Gianin (2002); if ϕ is also positively homogeneous, ρ is a coherent risk measure (Artzner et al., 1999).

PROPOSITION 2—Penalised and constrained values as risk functionals: *Let $\lambda > 0$ and $\bar{I} \geq 0$.*

- (i) *As functionals of u , Φ_λ and $\Psi_{\bar{I}}$ are monotone, convex and cash-additive, with $\Phi_\lambda(0) = \Psi_{\bar{I}}(0) = 0$, so $x \mapsto \Phi_\lambda(-x)$ and $x \mapsto \Psi_{\bar{I}}(-x)$ are convex risk measures.*
- (ii) *$\Psi_{\bar{I}}$ is positively homogeneous, so $x \mapsto \Psi_{\bar{I}}(-x)$ is a coherent risk measure. $\Phi_\lambda(ku) = k\Phi_{\lambda/k}(u)$ for $k > 0$, and $\Phi_\lambda(ku) = k\Phi_\lambda(u)$ holds for every $k > 0$ if and only if information has no value in the decision problem u .*
- (iii) *$\Psi_{\bar{I}}(u) = \inf_{\lambda > 0} \{\Phi_\lambda(u) + \lambda\bar{I}\}$.*

Part (i) for the penalised value is close to the representation of rationally inattentive preferences over menus by De Oliveira et al. (2017), in which the cost of information enters additively in the same way. Part (ii) restates the comparison across stakes of Behringer (2026, Section 6) in the language of risk measures: positive homogeneity of the constrained value is the invariance of the constrained solution to the stakes (Azrieli, 2021, Proposition 6), and its failure for the penalised value is the fall of the effective price from λ to λ/k .

Part (iii) identifies the duality between the two formulations with a known pair of risk measures. Fix a state and replace the endogenous action marginal by a fixed reference distribution P . The penalised problem $\max_r \{\sum_a r(a)u(a) - \lambda D(r\|P)\}$ then has the value $\lambda \ln E_P e^{u/\lambda}$, the entropic certainty equivalent of the payoff $u(A)$ with $A \sim P$, and the constrained problem $\max\{\sum_a r(a)u(a) : D(r\|P) \leq \bar{I}\}$ is the entropic value-at-risk of $u(A)$ at level $e^{-\bar{I}}$ (Ahmadi-Javid, 2012), whose dual formula $\inf_{\lambda>0} \{\lambda \ln E_P e^{u/\lambda} + \lambda \bar{I}\}$ is part (iii) in this special case. Mutual information adds the optimisation over the reference and the averaging over states, so the RI values are of entropic and entropic value-at-risk type rather than these functionals themselves. The hard constraint thus corresponds to a coherent risk measure defined by a bound on relative entropy, the linear penalty to a convex one defined by a penalty on it, and the shadow price of capacity is the minimising λ in (iii) (Behringer, 2026, Theorem 3).

REMARK—Relation to the risk quadrangle: For a single state and a fixed action marginal P , the constrained value $\inf_{\lambda>0} \lambda \{\bar{I} + \ln E_P e^{u/\lambda}\}$ is the risk measure of the quadrangle based on the Kullback–Leibler divergence in Grechuk et al. (2026, Example 10.1), with $\beta = \bar{I}$, and the penalised value is the entropic risk of its parent quadrangle. The values Φ_λ and $\Psi_{\bar{I}}$ lie outside that framework for two reasons. The reference measure is not given but chosen, as the action marginal, and the capacity is shared across states, so that a single price λ applies to all of them. The information premium of Section 4.2 is non-negative and, when information has value, positive after a sufficient rescaling of the payoffs, as for the subregular measures of Grechuk et al. (2026); it is not, however, a subregular deviation, because it is not convex (Proposition 3(v)) and vanishes at every scale in a decision problem with a dominant action.

REMARK—Monetary payoffs: If payoffs are monetary, $x \in \mathbb{R}^{\mathcal{A} \times \Omega}$, and evaluated with an increasing utility v , the value $\Phi_\lambda(v \circ x)$ is monotone in x , and convex in x when v is convex. It is cash-additive in money for every decision problem if and only if $v(z) = z + b$ with a single action and a single state, $\Phi_\lambda(v \circ x) = v(x)$. With $v(z) = az + b$ and $a \neq 1$, a sum of money c adds ac , so cash additivity holds only after the slope is normalised, and if v is strictly concave, Φ_λ is strictly concave on riskless payoffs and so is not convex. The same holds for $\Psi_{\bar{I}}$.

4.2. The information premium

DEFINITION 2—Information premium: The *information premium* of the decision problem u at price $\lambda > 0$ is

$$\mathcal{R}_\lambda^{\text{VoI}}(u) := \Phi_\lambda(u) - U_u(0) = \sup_{\bar{I} \geq 0} \{V_u(\bar{I}) - \lambda \bar{I}\}.$$

Geometrically, $\mathcal{R}_\lambda^{\text{VoI}}(u)$ is the vertical intercept of the supporting line of slope λ to the VoI. The subtracted term $U_u(0) = \Psi_0(u)$ is constant in λ but not in u : as a maximum of linear functions of u it is itself convex. The premium is thus the difference $\Phi_\lambda - \Psi_0$ of two convex functionals, which is in general neither convex nor concave (Proposition 3(v)).

PROPOSITION 3—Information premium: *Let $\lambda > 0$.*

- (i) $\mathcal{R}_\lambda^{\text{VoI}}(u) \geq 0$, *it is convex and non-increasing in λ , and it is zero if and only if $\lambda \geq V'_u(0+)$, the slope of the VoI at zero.*
- (ii) $\mathcal{R}_\lambda^{\text{VoI}}(u + c) = \mathcal{R}_\lambda^{\text{VoI}}(u)$ *for every constant c .*
- (iii) $\mathcal{R}_\lambda^{\text{VoI}}(ku) = k \mathcal{R}_{\lambda/k}^{\text{VoI}}(u)$ *for $k > 0$; hence $\mathcal{R}_\lambda^{\text{VoI}}(ku) \geq k \mathcal{R}_\lambda^{\text{VoI}}(u)$ for $k \geq 1$.*
- (iv) *If $V_u \geq V_w$ on $[0, \infty)$, then $\mathcal{R}_\lambda^{\text{VoI}}(u) \geq \mathcal{R}_\lambda^{\text{VoI}}(w)$ for every $\lambda > 0$, with strict inequality whenever $V_u > V_w$ on $(0, \infty)$ and $\mathcal{R}_\lambda^{\text{VoI}}(w) > 0$.*
- (v) $\mathcal{R}_\lambda^{\text{VoI}}$ *is in general neither monotone nor convex in u .*

The premium is therefore not a risk measure: by (ii) it ignores the level of payoffs, and by (v) it can fall when payoffs rise, because a payoff increase that makes one action dominant removes the value of information. It measures how much the decision problem rewards information. By (iii), raising the stakes is equivalent to lowering the price of information, so the premium grows more than proportionately with the stakes. This is the scale effect of [Arrow \(1996\)](#) at a given price of information; under a fixed capacity the VoI scales proportionately, $V_{ku} = kV_u$, and only its shadow price grows ([Behringer, 2026](#), Proposition 4).

5. APPLICATIONS

5.1. Diagnostic testing

A physician faces a binary diagnostic problem ([Stratonovich, 1975/2020](#), Section 10.1): a patient has a disease (θ_1) with prior probability p or not (θ_2), and the physician treats (a_1) or does not (a_2). A test with sensitivity $\text{se} = P(S = 1 \mid \theta_1)$ and specificity $\text{sp} = P(S = 0 \mid$

θ_2) costs $c > 0$ per patient. Take the identity utility matrix as a symmetric benchmark, with utility 1 for a correct decision and 0 otherwise; for arbitrary matrices see [Behringer and Belavkin \(2026\)](#). The VoI then depends on the prevalence: its maximum, the value of full information, is $1 - \max\{p, 1 - p\}$. At $p = \frac{1}{2}$ the mutual information of the test is

$$I(\Theta, S) = H\left(\frac{\text{se} + 1 - \text{sp}}{2}\right) - \frac{1}{2}H(\text{se}) - \frac{1}{2}H(\text{sp}), \quad (2)$$

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

The VoI $V(\bar{I})$ is the value of the best channel with capacity $\bar{I}$, so a given test with $I(\Theta, S) = \bar{I}$ is worth at most $V(\bar{I})$. At $p = \frac{1}{2}$ the bound is attained by symmetric tests, $\text{se} = \text{sp}$, which are binary symmetric channels; an asymmetric test with the same mutual information is worth less. The condition $V(I(\Theta, S)) \geq c$ is therefore necessary for ordering a test and sufficient for a symmetric one. Since V is continuous and strictly increasing up to the value of full information, for $0 < c < \frac{1}{2}$ there is a unique threshold I^* with $V(I^*) = c$. By concavity of V , the value $V(I_1) - V(I_0)$ of raising informativeness from I_0 to $I_1 = I_0 + \Delta$ is non-increasing in I_0 : a given gain in precision is worth more when baseline accuracy is low. With asymmetric utilities, capturing different costs of false positives and false negatives, the threshold depends on the prior, the utility matrix, se , sp and c .

5.2. Portfolio choice under tail risk

An investor with CARA utility $-e^{-\gamma w}$, $\gamma = 2$, allocates between a riskless asset ($r_f = 2\%$) and a risky asset. With initial wealth normalised to zero, terminal wealth is $w = r_f + \alpha R$, where R is the excess return of the risky asset. Two discrete distributions of R have the same mean (8%) and standard deviation (18%) but different higher moments (Table I); the skewed distribution has skewness -1.54 and excess kurtosis 2.49 and contains a crash state. The action is binary, $\alpha \in \{0, 1\}$, full investment or full allocation to the riskless asset, which keeps the optimal channel explicit while capturing the decision whether to be in or out of equities. The VoI is reported as a percentage of the utility level $|u_0| = e^{-\gamma r_f}$ of the riskless position, and the shadow price λ^* in the same units, as the slope per nat of $V/|u_0|$; Appendix B derives all numbers.

At $\bar{I} = 0.30$ bits the VoI is 5.88% of $|u_0|$ under the symmetric distribution and 10.86% under the skewed one, 85% more. The Gibbs channel (1) explains why. The crash state

TABLE I

EXCESS RETURN DISTRIBUTIONS WITH IDENTICAL MEANS AND STANDARD DEVIATIONS

Distribution	States	Excess returns	Probabilities
Symmetric	Loss, Gain	−10%, +26%	50%, 50%
Skewed	Crash, Normal, Gain	−40%, +8.8%, +30%	10%, 70.75%, 19.25%

TABLE II

VOI AND SHADOW PRICE PER NAT AT FIXED CAPACITIES ($\gamma = 2$, $r_f = 2\%$, $\alpha \in \{0, 1\}$)

$\bar{I}$ (bits)	VoI Sym (%)	VoI Skew (%)	λ_{Sym}^*	λ_{Skew}^*	VoI ratio
0.10	2.64	5.90	0.303	0.551	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

(−40%) accounts for 81.5% of the mutual information, $\mu(\theta)D(q(\cdot | \theta) \| P)$ summed over states, despite a prior probability of 10% (Table III). The utility gap between exiting and staying invested in the crash state, divided by the shadow price, is about 5.6, so the Gibbs kernel multiplies the unconditional odds of exit, $P(\text{exit})/P(\text{stay}) \approx 0.163/0.837$, by $e^{5.6} \approx 277$, and the exit signal arrives with probability 0.98 in a crash. The concentration is driven by the payoff gap, not by the prior probability.

TABLE III

SHARES OF MUTUAL INFORMATION 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%
$P(\text{exit signal} \text{downside state})$	63.4%	98.2%

A model that summarises returns by their mean and variance assigns the same value of information to both distributions. In this example the skewed distribution makes information 85 to 124% more valuable, depending on the capacity, and the gap is largest when capacity is tightest. Models that value data by symmetric benchmarks therefore underprice tail-sensitive information. By Proposition 3(iv), wherever the VoI of the skewed problem exceeds that of the symmetric one at every positive capacity, its information premium is at least as large at every price.

Table IV shows that the VoI more than doubles when γ doubles, more strongly for the skewed distribution. A change in risk aversion is not a rescaling of the payoff matrix, so this is not an instance of Proposition 3(iii), but it has the same source: higher risk aversion widens the utility gap in the crash state, and the Gibbs kernel weights that gap exponentially.

TABLE IV
VOI AND 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

5.3. Climate research under deep uncertainty

Key parameters of climate policy, such as climate sensitivity and the damage function, are subject to deep expert disagreement (Heal and Millner, 2014). The question here is how limited research capacity should be allocated across states and what an additional unit of capacity is worth. A policymaker chooses a mitigation level a given uncertain climate sensitivity Θ and payoff $u(a, \Theta)$. By Proposition 1, the penalised value is $\max_P \sum_{\theta} \mu(\theta) W_{\lambda}(\theta; P)$: the outer expectation is over states under the prior, and the inner certainty equivalent is over mitigation levels drawn from the unconditional distribution P chosen by the policymaker. As $\lambda \rightarrow \infty$ the Gibbs channel collapses to P and the policymaker acts on the prior; as $\lambda \rightarrow 0$ the channel approaches full revelation and

$\Phi_\lambda(u) \rightarrow \sum_\theta \mu(\theta) \max_a u(a, \theta)$. Attitudes to damage risk are carried by the curvature of u , and λ prices capacity.

By (1), the channel weights each action in state θ by $e^{u(a, \theta)/\lambda}$, so capacity concentrates on the states in which the payoff gap between mitigation levels is largest. High-sensitivity, high-damage states therefore absorb a share of research capacity far above their prior probability, as the crash state does in Section 5.2. An observer of such a research programme might infer a pessimistic distortion of beliefs or an ambiguity-averse objective, but no such distortion is present: the prior is untouched, and the allocation follows from the payoff gaps under a binding capacity.

Where research capacity is fixed by institutional and technological limits over the horizon considered, the constrained formulation is the appropriate one (Behringer, 2026, Section 6). Its shadow price $\lambda^*(\bar{I}) = V'(\bar{I})$ is the marginal value of expanding capacity, in damage units per nat, which gives a decision-theoretic measure of the value of funding climate science. Because plausible damage functions differ by large factors and their tails are poorly identified (Pindyck, 2013), the response to a rescaling of damages matters. At a fixed capacity the VoI scales proportionately, $V_{ku}(\bar{I}) = kV_u(\bar{I})$, and so does the shadow price. At a fixed price of information, Proposition 3(iii) gives $\mathcal{R}_\lambda^{\text{VoI}}(ku) = k\mathcal{R}_{\lambda/k}^{\text{VoI}}(u) \geq k\mathcal{R}_\lambda^{\text{VoI}}(u)$ for $k \geq 1$: doubling the stakes at least doubles the premium, because raising the stakes lowers the effective price of information. A worsening climate outlook is therefore an argument for expanding research as well as mitigation.

6. CONCLUSION

Read as functionals of the payoffs, the two formulations of rational inattention are risk measures of a familiar kind. The penalised value is an expected entropic certainty equivalent whose reflection is a convex risk measure; the constrained value is positively homogeneous, of entropic value-at-risk type, and its reflection is coherent; and the conjugate duality between them is the dual formula of the entropic value-at-risk. The risk attitude in these formulas is an option value: the informed agent evaluates the payoffs of the actions she may choose as a risk seeker would, because information lets her select among them. The information premium, which is not a risk measure, orders decision problems by their value of information and grows more than proportionately with the stakes at a given price.

In the applications, a binding capacity directs information to states with large payoff gaps, so that tail states receive more attention than their probability alone would warrant.

APPENDIX A: PROOFS

PROOF OF PROPOSITION 1: Mutual information is the smallest average relative entropy of the conditional action distributions with respect to a common reference, $I(q) = \min_P \sum_{\theta} \mu(\theta) D(q(\cdot | \theta) \| P)$, attained at the action marginal of q (Cover and Thomas, 2006, Section 10.8). Interchanging the two maximisations, for fixed P the problem separates across states, and in each state the variational formula for relative entropy (Donsker and Varadhan, 1975), $\max_r \{ \sum_a r(a) u(a, \theta) - \lambda D(r \| P) \} = \lambda \ln \sum_a P(a) e^{u(a, \theta)/\lambda}$, is attained at the Gibbs kernel (1). At a maximising P , the first-order condition of the outer problem, $\sum_{\theta} \mu(\theta) e^{u(a, \theta)/\lambda} / \sum_b P(b) e^{u(b, \theta)/\lambda} \leq 1$ with equality where $P(a) > 0$, states that the action marginal of (1) is P . *Q.E.D.*

PROOF OF PROPOSITION 2: (i) Both values are maxima over channels of $\langle q, u \rangle$ less a term that does not depend on u , hence maxima of affine functions of u and convex. Monotonicity holds because $q \geq 0$, and cash additivity because $\sum_a q(a | \theta) = 1$, so that $\langle q, u + c \rangle = \langle q, u \rangle + c$. At $u = 0$ the uninformative channel is optimal in both problems, with value 0.

(ii) The feasible set of $\Psi_{\bar{I}}$ does not depend on u , so $\Psi_{\bar{I}}(ku) = k \Psi_{\bar{I}}(u)$. For the penalised value, $\langle q, ku \rangle - \lambda I(q) = k [\langle q, u \rangle - (\lambda/k) I(q)]$. By Section 2, $\Phi_{\lambda'}(u) = U_u(0) + \sup_{\bar{I}} \{ V_u(\bar{I}) - \lambda' \bar{I} \}$, which is non-increasing in λ' , tends to $U_u(0) + \max V_u$ as $\lambda' \rightarrow 0$, and tends to $U_u(0)$ as $\lambda' \rightarrow \infty$ because V_u is bounded and continuous at 0 with $V_u(0) = 0$. It is therefore constant in λ' if and only if $\max V_u = 0$, and $\Phi_{\lambda}(ku) = k \Phi_{\lambda}(u)$ for every $k > 0$ requires exactly this.

(iii) This is the biconjugate identity of Behringer (2026, Theorem 2), stated there with $\lambda \geq 0$. Excluding $\lambda = 0$ does not change the infimum: since $0 \leq I(q) \leq \ln |\mathcal{A}|$, $|\Phi_{\lambda}(u) - \Phi_0(u)| \leq \lambda \ln |\mathcal{A}|$, so $\Phi_{\lambda}(u) + \lambda \bar{I} \rightarrow \Phi_0(u)$ as $\lambda \downarrow 0$. *Q.E.D.*

PROOF OF PROPOSITION 3: (i) The supremum includes $\bar{I} = 0$, where the objective is 0. As a supremum of affine functions of λ with slopes $-\bar{I} \leq 0$, the premium is convex and non-increasing in λ . It is zero if and only if $V_u(\bar{I}) \leq \lambda \bar{I}$ for all $\bar{I}$, that is,

$\lambda \geq \sup_{\bar{I} > 0} V_u(\bar{I})/\bar{I}$; by concavity and $V_u(0) = 0$ the ratio is non-increasing in $\bar{I}$, so the supremum is $V'_u(0+)$.

(ii) Adding c to every payoff adds c to $\Phi_\lambda(u)$ by Proposition 2(i) and to $U_u(0)$.

(iii) $U_{ku}(0) = kU_u(0)$ and $\Phi_\lambda(ku) = k\Phi_{\lambda/k}(u)$ by Proposition 2(ii). For $k \geq 1$, $\lambda/k \leq \lambda$, and the premium is non-increasing in the price by (i).

(iv) $V_u \geq V_w$ gives $V_u(\bar{I}) - \lambda\bar{I} \geq V_w(\bar{I}) - \lambda\bar{I}$ for every $\bar{I}$. If $\mathcal{R}_\lambda^{\text{VoI}}(w) > 0$, the supremum for w is attained at some $\bar{I}_w > 0$, since V_w is continuous and $V_w(\bar{I}) - \lambda\bar{I} \rightarrow -\infty$; then $\mathcal{R}_\lambda^{\text{VoI}}(u) \geq V_u(\bar{I}_w) - \lambda\bar{I}_w > V_w(\bar{I}_w) - \lambda\bar{I}_w = \mathcal{R}_\lambda^{\text{VoI}}(w)$.

(v) Let $\Omega = \{\theta_1, \theta_2\}$ with μ uniform, $\mathcal{A} = \{a_1, a_2\}$, and $u(a_i, \theta_j) = 1$ if $i = j$ and 0 otherwise. Information has value in u , so $\mathcal{R}_\lambda^{\text{VoI}}(u) > 0$ for $\lambda < V'_u(0+)$ by (i). For $M \geq 2$, let $w_1 = 2u + M\mathbf{1}_{a=a_1}$ and $w_2 = -M\mathbf{1}_{a=a_1}$, so that $u = \frac{1}{2}(w_1 + w_2)$. In w_1 action a_1 is optimal in both states and in w_2 action a_2 is, so information has no value and both premia are zero. Hence $\mathcal{R}_\lambda^{\text{VoI}}(\frac{1}{2}w_1 + \frac{1}{2}w_2) > \frac{1}{2}\mathcal{R}_\lambda^{\text{VoI}}(w_1) + \frac{1}{2}\mathcal{R}_\lambda^{\text{VoI}}(w_2)$, and convexity fails. Since $w_1 \geq 2u \geq u$ while $\mathcal{R}_\lambda^{\text{VoI}}(w_1) = 0 < \mathcal{R}_\lambda^{\text{VoI}}(u)$, monotonicity fails as well. *Q.E.D.*

APPENDIX B: ANALYTICAL DERIVATION OF THE PORTFOLIO APPLICATION

This appendix derives all quantitative results of the portfolio application (Section 5.2) from first principles. Both distributions have finite support and the action space is binary, so the Gibbs optimal signal reduces to an explicit algebraic fixed-point system.

The fixed-point equation is a polynomial in the marginal exit-signal probability q ; factoring out the trivial root $q = 1$ (the uninformative signal, Section B.4) reduces it to a *closed-form* expression for the symmetric case and a *quadratic* for the skewed case. A single bisection on the shadow price λ then pins mutual information to the target $\bar{I}$. No simulation is involved.

Notation follows Section 2: μ is the prior and λ the price of information per nat, in the utility units that the Gibbs weights $e^{u/\lambda}$ require; Table II reports it divided by $|u_0|$, in the units of its VoI columns. Signals are recommendations: s_{exit} recommends $\alpha = 0$ and s_{stay} recommends $\alpha = 1$, so that $p^*(s | \theta)$ is the channel $q(\alpha | \theta)$ of (1) and q is its marginal probability $P(0)$ of exit.

B.1. *Primitives*

The investor has CARA utility $U(w) = -e^{-\gamma w}$ with $\gamma = 2$ and riskless return $r_f = 0.02$. Initial wealth is normalised to zero. The binary action space is $\mathcal{A} = \{0, 1\}$, where $\alpha = 1$ denotes full investment in the risky asset and $\alpha = 0$ the riskless asset. The realised payoff under action α in a state with excess return $R(\theta)$ is $w = r_f + \alpha R(\theta)$, so:

$$U(\alpha, \theta) = -\exp(-\gamma[r_f + \alpha R(\theta)]). \quad (3)$$

Common riskless-action utility. Since $U(0, \theta)$ is state-independent:

$$u_0 \equiv U(0, \theta) = -e^{-\gamma r_f} = -e^{-0.04} \approx -0.96079 \quad \text{for all } \theta. \quad (4)$$

Symmetric distribution. States θ_1 and θ_2 with prior $\mu = (\frac{1}{2}, \frac{1}{2})$ and excess returns $R_1 = -0.10$, $R_2 = +0.26$:

$$u_1^{\text{sym}} = U(1, \theta_1) = -e^{-2(r_f + R_1)} = -e^{-2(-0.08)} = -e^{+0.16} \approx -1.17351, \quad (5)$$

$$u_2^{\text{sym}} = U(1, \theta_2) = -e^{-2(r_f + R_2)} = -e^{-2(0.28)} = -e^{-0.56} \approx -0.57121. \quad (6)$$

Skewed distribution. States $\theta_1, \theta_2, \theta_3$ with $\mu = (0.10, 0.7075, 0.1925)$ and excess returns $R_1 = -0.40$, $R_2 = +0.088$, $R_3 = +0.30$:

$$u_1^{\text{sk}} = -e^{-2(r_f + R_1)} = -e^{-2(-0.38)} = -e^{+0.76} \approx -2.13828, \quad (7)$$

$$u_2^{\text{sk}} = -e^{-2(r_f + R_2)} = -e^{-2(0.108)} = -e^{-0.216} \approx -0.80574, \quad (8)$$

$$u_3^{\text{sk}} = -e^{-2(r_f + R_3)} = -e^{-2(0.32)} = -e^{-0.64} \approx -0.52729. \quad (9)$$

Moment verification. Both distributions of the excess return share mean 8% and standard deviation 18%. For the skewed distribution,

$$E\{R\} = 0.10(-0.40) + 0.7075(0.088) + 0.1925(0.30) = 0.080,$$

$$\text{Var}(R) = 0.10(0.48)^2 + 0.7075(0.008)^2 + 0.1925(0.22)^2 = 0.0324,$$

so the standard deviation is 0.18. Skewness equals -1.54 and excess kurtosis equals 2.49 .

B.2. Prior-optimal actions

The switch criterion for CARA utility under action $\alpha = 1$ versus $\alpha = 0$ is $U(1, \theta) \leq U(0, \theta)$, which simplifies to $R(\theta) \leq 0$. The agent prefers to invest whenever the excess return is positive.

Symmetric.

$$E\{U(\alpha=1)\} = \frac{1}{2}(-1.17351) + \frac{1}{2}(-0.57121) = -0.87236, \quad (10)$$

$$E\{U(\alpha=0)\} = -0.96079. \quad (11)$$

Since $-0.87236 > -0.96079$, the prior-optimal action is invest:

$$U^{\text{sym}}(0) = -0.87236. \quad (12)$$

Skewed.

$$\begin{aligned} E\{U(\alpha=1)\} &= 0.10(-2.13828) + 0.7075(-0.80574) \\ &\quad + 0.1925(-0.52729) = -0.88539, \end{aligned} \quad (13)$$

$$E\{U(\alpha=0)\} = -0.96079. \quad (14)$$

Prior-optimal action is again invest:

$$U^{\text{sk}}(0) = -0.88539. \quad (15)$$

B.3. First-best (perfect information)

With perfect information the investor observes θ and exits whenever $R(\theta) < 0$.

Symmetric. Exit in state θ_1 ($R = -0.10$), invest in θ_2 ($R = +0.26$):

$$U_{\infty}^{\text{sym}} = \frac{1}{2}(-0.96079) + \frac{1}{2}(-0.57121) = -0.76600. \quad (16)$$

Skewed. Exit in crash state θ_1 ($R = -0.40$), invest in θ_2 and θ_3 :

$$U_{\infty}^{\text{sk}} = 0.10(-0.96079) + 0.7075(-0.80574) + 0.1925(-0.52729) = -0.76764. \quad (17)$$

B.4. Gibbs optimal signal: analytical structure

Binary signal alphabet. We use a signal alphabet $\mathcal{S} = \{s_{\text{exit}}, s_{\text{stay}}\}$ with induced optimal actions $a^*(s_{\text{exit}}) = 0$ and $a^*(s_{\text{stay}}) = 1$. Let $q = p(s_{\text{exit}})$ be the marginal probability of the exit signal. By Proposition 1, the Gibbs optimal conditional distribution at price $\lambda > 0$ is:

$$p^*(s | \theta_i) = \frac{p(s) e^{u(s, \theta_i)/\lambda}}{Z(\theta_i)}, \quad (18)$$

where $u(s_{\text{exit}}, \theta_i) = u_0$ for all i (exit action yields u_0 regardless of state) and $u(s_{\text{stay}}, \theta_i) = u_i$ (invest-action utility in state i). Defining:

$$A_0 = e^{u_0/\lambda}, \quad A_i = e^{u_i/\lambda}, \quad (19)$$

the partition functions are:

$$Z(\theta_i) = qA_0 + (1 - q)A_i. \quad (20)$$

Free-energy identity and the value of information. Before deriving the fixed-point equation for q , we record the key formula for $U(\bar{I})$ and prove the free-energy identity underlying the primal–dual identity (Behringer, 2026, Theorem 2), following Stratonovich’s (1975/2020, Section 9.4) representation. At the Gibbs optimum $p^*(s | \theta) = p(s)e^{u/\lambda}/Z(\theta)$, substituting into $E\{U\} - \lambda I$ gives:

$$\begin{aligned} E\{U\} - \lambda I &= \sum_{\theta} \mu(\theta) \sum_s p^*(s | \theta) \left[u(s, \theta) - \lambda \log \frac{p^*(s | \theta)}{p(s)} \right] \\ &= \sum_{\theta} \mu(\theta) \sum_s p^*(s | \theta) [u(s, \theta) - u(s, \theta) + \lambda \log Z(\theta)] \\ &= \sum_{\theta} \mu(\theta) \lambda \log Z(\theta) = E_{\theta}\{F(\theta)\}, \end{aligned} \quad (21)$$

where $F(\theta) = \lambda \log Z(\theta)$ is the free energy. The primal–dual identity then follows because Legendre–Fenchel duality gives $U(\bar{I}) = \min_{\lambda \geq 0} \{E_{\theta}\{F(\theta)\} + \lambda \bar{I}\}$, and at the interior optimum λ^* where the constraint binds, this minimum is attained:

$$U(\bar{I}) = E_{\theta}\{F(\theta)\} \big|_{\lambda=\lambda^*} + \lambda^* \bar{I}, \quad (22)$$

with $\bar{I}$ measured in nats (a budget of b bits is $b \ln 2$ nats), so that λ^* is the price per nat. The Stratonovich/Shannon VoI (Section 2) follows directly:

$$V(\bar{I}) = U(\bar{I}) - U(0) = E_\theta\{F(\theta)\}\big|_{\lambda=\lambda^*} + \lambda^* \bar{I} - U(0). \quad (23)$$

Marginal self-consistency and explicit roots for q^* . The marginal $p(s_{\text{exit}}) = q$ must equal the mixture $q = \sum_i \mu_i p^*(s_{\text{exit}} | \theta_i)$.¹ From (18) the Gibbs optimal conditional for the exit signal is $p^*(s_{\text{exit}} | \theta_i) = q A_0 / Z(\theta_i)$; substituting and dividing by $q > 0$ gives the general self-consistency equation:

$$1 = A_0 \sum_{i=1}^n \frac{\mu_i}{Z(\theta_i)} = A_0 \sum_{i=1}^n \frac{\mu_i}{q A_0 + (1-q) A_i}. \quad (24)$$

The trivial root $q = 1$. Equation (24) always admits $q = 1$ as a solution: when $q = 1$ all partition functions collapse to $Z(\theta_i) = A_0$, so the right-hand side equals $A_0 \sum_i \mu_i / A_0 = 1$ regardless of the utility values. Economically, $q = 1$ corresponds to an uninformative signal that always fires the exit action; it is the trivial solution outside the interior of the information set.

Factoring out the trivial root. Cross-multiplying (24) by $\prod_j Z(\theta_j)$ yields a degree- n polynomial in q with $q = 1$ as a root. At $q = 1$, we have $p^*(s_{\text{exit}} | \theta) = 1$ for all θ , so the signal is degenerate and carries no information. Consequently, the mutual information vanishes: $I(\Theta, S) = 0$. Factoring out $(q - 1)$ reduces the polynomial degree by one, and the unique interior root $q^* \in (0, 1)$ is identified from the reduced polynomial.

Symmetric case ($n = 2$, $\mu_1 = \mu_2 = \frac{1}{2}$). Cross-multiplying (24) by $2Z_1 Z_2$ gives:

$$2Z_1 Z_2 = A_0(Z_1 + Z_2). \quad (25)$$

¹Economically, q is the unconditional probability that the investor receives an exit signal, before learning the state. If the optimal conditional signal probabilities $p^*(s | \theta_i)$ implied a different unconditional exit probability than q , the signal design would be self-contradictory: the marginal distribution of signals facing the investor would not match the one the optimal signal was designed around. The self-consistency condition therefore requires q to be a fixed point: the assumed marginal q and the prior-weighted average $\sum_i \mu_i p^*(s_{\text{exit}} | \theta_i)$ must agree, as for the action marginal in Proposition 1.

Expanding with $Z_i = qA_0 + (1 - q)A_i$ and collecting by powers of q :

$$2(A_0 - A_1)(A_0 - A_2)q^2 + [-2A_0^2 + 3A_0(A_1 + A_2) - 4A_1A_2]q + [2A_1A_2 - A_0(A_1 + A_2)] = 0. \quad (26)$$

Since both $q = 1$ and $q = q^*$ are the roots, Vieta's product-of-roots formula immediately gives the closed-form interior root:

$$q^* = \frac{2A_1A_2 - A_0(A_1 + A_2)}{2(A_0 - A_1)(A_0 - A_2)}. \quad (27)$$

Skewed case ($n = 3$). Cross-multiplying (24) by $Z_1Z_2Z_3$ and dividing by A_0 gives $f(q) = 0$ for the cubic

$$f(q) := A_0^{-1}Z_1(q)Z_2(q)Z_3(q) - \sum_{i=1}^3 \mu_i \prod_{j \neq i} Z_j(q). \quad (28)$$

Again $q = 1$ is a root. Dividing out $(q - 1)$ yields the reduced quadratic $h(q) = \kappa_2q^2 + \kappa_1q + \kappa_0$, so that $f(q) = (q - 1)h(q)$, with coefficients:

$$\kappa_2 = \frac{(A_0 - A_1)(A_0 - A_2)(A_0 - A_3)}{A_0}, \quad (29)$$

$$\kappa_0 = \mu_1A_2A_3 + \mu_2A_1A_3 + \mu_3A_1A_2 - \frac{A_1A_2A_3}{A_0}, \quad (30)$$

$$\kappa_1 = A_0 \left(A_0 - \sum_{i=1}^3 \mu_i A_i \right) - \kappa_2 - \kappa_0. \quad (31)$$

Here κ_2 is the leading coefficient of the cubic (28), $\kappa_0 = h(0) = -f(0)$ follows from evaluating the cubic at $q = 0$, and κ_1 is determined by $h(1) = \kappa_2 + \kappa_1 + \kappa_0 = A_0(A_0 - \sum_i \mu_i A_i)$, which is obtained by differentiating the cubic at $q = 1$, since $f'(1) = h(1)$.² The unique root in $(0, 1)$ is then³

$$q^* = \frac{-\kappa_1 - \sqrt{\kappa_1^2 - 4\kappa_2\kappa_0}}{2\kappa_2}. \quad (32)$$

²At $q = 1$ all $Z_i = A_0$, so $f'(1) = A_0 \sum_i (A_0 - A_i) - A_0 \sum_i (1 - \mu_i)(A_0 - A_i) = A_0 \sum_i \mu_i (A_0 - A_i) = A_0(A_0 - \bar{A})$, where $\bar{A} = \sum_i \mu_i A_i$.

³Here $\kappa_2 > 0$, and the negative sign before the square root selects the root in $(0, 1)$; the other root of h satisfies $q > 1$ and is discarded.

Numerical evaluation for the skewed distribution at $\bar{I} = 0.30$ is carried out in Section B.5.

Bisection for λ^* and the explicit mutual information formula. For fixed target $\bar{I}$ (in nats), the shadow price λ^* solves:

$$I(\Theta, S; \lambda^*) = \bar{I}. \quad (33)$$

The mutual information of the Gibbs channel at price λ is given in (34) below and is strictly decreasing in λ , as the numerical evaluations in Section B.5 confirm.

$$I(\Theta, S; \lambda) = \sum_{i=1}^n \mu_i \left[\frac{qA_0}{Z_i} \ln \frac{A_0}{Z_i} + \frac{(1-q)A_i}{Z_i} \ln \frac{A_i}{Z_i} \right], \quad (34)$$

where $q = q^*(\lambda)$ satisfies (24) at price λ , $A_0 = e^{u_0/\lambda}$, $A_i = e^{u_i/\lambda}$, and $Z_i = qA_0 + (1-q)A_i$.

At the Gibbs optimum, the free-energy identity (21) gives $E\{U\} - \lambda I = E_\theta\{F(\theta)\}$, so (34) can equivalently be written as:

$$I(\Theta, S; \lambda) = \frac{E\{U\} - E_\theta\{F(\theta)\}}{\lambda}. \quad (35)$$

For the bisection, (33) and (34) are employed; (35) is useful for the soft constraint problem where λ is given and $E\{U\} - \lambda I$ is recovered directly from the free energy.

The function $I(\Theta, S; \lambda)$ is strictly decreasing in λ : a higher price flattens the Gibbs weights, spreading $p^*(s | \theta)$ more uniformly and reducing information transmission. It satisfies $I \rightarrow \bar{I}_{\text{sat}}$ (the saturation budget, see Section B.7) as $\lambda \rightarrow 0$ and $I \rightarrow 0$ as $\lambda \rightarrow \infty$. Bisection on the monotone residual $I(\lambda) - \bar{I}$ therefore converges to arbitrary precision: initialise a bracket $[\lambda_{\text{lo}}, \lambda_{\text{hi}}]$ with $I(\lambda_{\text{lo}}) > \bar{I} > I(\lambda_{\text{hi}})$, set $\lambda_{\text{mid}} = (\lambda_{\text{lo}} + \lambda_{\text{hi}})/2$, and replace the endpoint with the same sign as the residual by λ_{mid} . The bracket halves at each step, achieving 2^{-k} relative error after k iterations.

B.5. Numerical results at $\bar{I} = 0.30$ bits

Symmetric distribution: explicit q^* from the closed-form expression (27). At the bisected price $\lambda_{\text{sym}}^* \approx 0.1787$ in utility units, which is $0.1787/|u_0| = 0.186$ in the units of Table II:

$$A_0 = e^{u_0/\lambda^*} = e^{-0.96079/0.1787} \approx 0.004617,$$

$$A_1 = e^{u_1^{\text{sym}}/\lambda^*} = e^{-1.17351/0.1787} \approx 0.001404,$$

$$A_2 = e^{u_2^{\text{sym}}/\lambda^*} = e^{-0.57121/0.1787} \approx 0.040871.$$

Substituting into (27):

$$\text{Numerator: } 2A_1A_2 - A_0(A_1 + A_2)$$

$$= 2(0.001404)(0.040871) - (0.004617)(0.042275)$$

$$\approx 0.0001147 - 0.0001952 = -0.0000805,$$

$$\text{Denominator: } 2(A_0 - A_1)(A_0 - A_2)$$

$$= 2(0.003213)(-0.036254) \approx -0.0002330,$$

giving:

$$q^* = \frac{-0.0000805}{-0.0002330} \approx 0.345. \quad (36)$$

TABLE V

GIBBS OPTIMAL CONDITIONAL DISTRIBUTION, SYMMETRIC CASE ($\bar{I} = 0.30$ BITS)

State	$p^*(s_{\text{exit}} \theta)$	$p^*(s_{\text{stay}} \theta)$	ΔI_i (bits)	Share of I
θ_1 : loss ($R = -10\%$)	0.634	0.366	0.125	41.6%
θ_2 : gain ($R = +26\%$)	0.056	0.944	0.175	58.4%
Marginal	0.345	0.655	0.300	100%

Why the loss state receives 41.6% of capacity. The Gibbs weight on state θ_1 relative to θ_2 is governed by the utility difference: the gain state has invest utility $u_2^{\text{sym}} = -0.571$ versus the loss state's $u_1^{\text{sym}} = -1.174$. The exit signal (action $\alpha = 0$, utility $u_0 = -0.961$) lies between these two values, so the exit signal is most informative where the invest action is most harmful, in the loss state. Nevertheless, since both states have equal prior weight and the utility gap between invest and exit is moderate in both directions, the capacity concentrates slightly more on the gain state (58.4%) because discriminating the high-utility gain state from the exit option is also valuable.

Posterior probability of loss given exit signal. By Bayes' rule and the equal prior:

$$P(\theta_1 | s_{\text{exit}}) = \frac{0.5 \times 0.634}{0.5 \times 0.634 + 0.5 \times 0.056} = 0.919. \quad (37)$$

The conditional exit probability in the loss state is $p^*(s_{\text{exit}} | \theta_1) = 0.634 \approx 63.4\%$, matching Table III in the main text.

Value of information. Applying the primal–dual identity (22) with $\bar{I} = 0.30 \times \ln 2 = 0.2079$ nats and the partition functions evaluated at the bisected λ^* :

$$U^{\text{sym}}(0.30) = E_{\theta}\{\lambda^* \log Z(\theta)\} + \lambda^* \bar{I}, \quad (38)$$

$$V^{\text{sym}}(0.30) = U^{\text{sym}}(0.30) - U^{\text{sym}}(0) = 0.0565 \quad (\approx 5.88\% \text{ of } |u_0|). \quad (39)$$

The percentage is computed as $100 \times V(\bar{I})/|u_0|$, where $u_0 = -e^{-\gamma^* r f}$ is the riskless-asset utility, common to both distributions. It measures the VoI relative to the utility level of the riskless position, not as a share of wealth, which is normalised to zero.

Skewed distribution: reduced quadratic from (32). At the bisected price $\lambda_{\text{sk}}^* \approx 0.2094$ in utility units (0.218 in the units of Table II):

$$A_0 = e^{-0.96079/0.2094} \approx 0.01016, \quad A_1 = e^{-2.13828/0.2094} \approx 3.668 \times 10^{-5},$$

$$A_2 = e^{-0.80574/0.2094} \approx 0.02131, \quad A_3 = e^{-0.52729/0.2094} \approx 0.08058.$$

The reduced-quadratic coefficients from (29), (30) and (31) are:

$$\begin{aligned} \kappa_2 &= \frac{(A_0 - A_1)(A_0 - A_2)(A_0 - A_3)}{A_0} \\ &\approx \frac{(0.01012)(-0.01115)(-0.07041)}{0.01016} \approx 7.822 \times 10^{-4}, \end{aligned}$$

$$\begin{aligned} \kappa_0 &= \mu_1 A_2 A_3 + \mu_2 A_1 A_3 + \mu_3 A_1 A_2 - \frac{A_1 A_2 A_3}{A_0} \\ &\approx (0.10)(1.717 \times 10^{-3}) + (0.7075)(2.956 \times 10^{-6}) \\ &\quad + (0.1925)(7.817 \times 10^{-7}) - \frac{6.299 \times 10^{-8}}{0.01016} \\ &\approx 1.740 \times 10^{-4} - 6.198 \times 10^{-6} \approx 1.678 \times 10^{-4}, \end{aligned}$$

$$\begin{aligned}
\kappa_1 &= A_0(A_0 - \sum_i \mu_i A_i) - \kappa_2 - \kappa_0 \\
&= 0.01016(0.01016 - 0.03059) - 7.822 \times 10^{-4} - 1.678 \times 10^{-4} \\
&\approx -2.076 \times 10^{-4} - 9.500 \times 10^{-4} \approx -1.158 \times 10^{-3}.
\end{aligned}$$

Applying (32):

$$q^* = \frac{1.158 - \sqrt{1.158^2 - 4(0.7822)(0.1678)}}{2(0.7822)} \approx 0.163, \quad (40)$$

where all coefficients are expressed in units of 10^{-3} , and the other root of h is 1.317. Table VI records the full Gibbs conditional distribution.

TABLE VI
GIBBS OPTIMAL CONDITIONAL DISTRIBUTION, SKEWED CASE ($\bar{I} = 0.30$ BITS)

State	$p^*(s_{\text{exit}} \theta)$	$p^*(s_{\text{stay}} \theta)$	ΔI_i (bits)	Share of I
θ_1 : crash ($R = -40\%$)	0.982	0.018	0.244	81.5%
θ_2 : normal ($R = +8.8\%$)	0.085	0.915	0.027	8.9%
θ_3 : gain ($R = +30\%$)	0.024	0.976	0.029	9.6%
Marginal	0.163	0.837	0.300	100%

Why 81.5% of capacity concentrates on the crash state. The Gibbs weight on the exit signal in the crash state is governed by the utility ratio:

$$\frac{p^*(s_{\text{exit}} | \theta_1)}{p^*(s_{\text{stay}} | \theta_1)} = \frac{q^* e^{u_0/\lambda^*}}{(1 - q^*) e^{u_1^{\text{sk}}/\lambda^*}} = \frac{q^*}{1 - q^*} e^{(u_0 - u_1^{\text{sk}})/\lambda^*}. \quad (41)$$

The exponent is $(u_0 - u_1^{\text{sk}})/\lambda^* = (-0.9608 + 2.1383)/0.2094 = 1.1775/0.2094 \approx 5.62$, giving $e^{5.62} \approx 277$. The odds of exit in the crash state are therefore $(0.163/0.837) \times 277 \approx 53.9$, so that $p^*(s_{\text{exit}} | \theta_1) = 53.9/54.9 \approx 0.982$: the exit signal occurs with near certainty in the crash state. The corresponding state-level mutual information contribution (see Section B.6) accounts for 81.5% of the total 0.30-bit budget.

Posterior probability of crash given exit signal.

$$\begin{aligned} P(\theta_1 | s_{\text{exit}}) &= \frac{0.10 \times 0.982}{0.10 \times 0.982 + 0.7075 \times 0.085 + 0.1925 \times 0.024} \\ &= \frac{0.0982}{0.163} = 0.603. \end{aligned} \quad (42)$$

The exit signal raises the crash posterior from the 10% prior to 60.3%. Despite this, $P(s_{\text{exit}} | \theta_1) = 0.982 \approx 98.2\%$: the exit signal is triggered with near certainty when the crash occurs.

Value of information.

$$V^{\text{sk}}(0.30) = U^{\text{sk}}(0.30) - U^{\text{sk}}(0) = 0.1043 \quad (\approx 10.86\% \text{ of } |u_0|). \quad (43)$$

Ratio and its interpretation.

$$\frac{V^{\text{sk}}(0.30)}{V^{\text{sym}}(0.30)} = \frac{0.1043}{0.0565} \approx 1.845. \quad (44)$$

Both distributions have identical means and variances, so a model that summarises returns by their mean and variance assigns the same VoI to both. The 85% premium is attributable entirely to skewness and excess kurtosis entering the Gibbs partition function.

B.6. State-level mutual information decomposition

In the context of the portfolio application, ΔI_i measures the share of the investor's total information budget I that the optimally designed signal devotes to resolving uncertainty in state θ_i ; equivalently, how much attention capacity the rational investor allocates to detecting each possible return scenario. A large ΔI_i means the signal is highly informative about whether state i has occurred, while a small ΔI_i means that state is largely ignored by the optimal signal design.

For any signal distribution $p^*(s | \theta)$, the formal definition is:

$$\Delta I_i = \mu_i \sum_{s \in \mathcal{S}} p^*(s | \theta_i) \log \frac{p^*(s | \theta_i)}{p(s)}, \quad \sum_i \Delta I_i = \bar{I}. \quad (45)$$

For the skewed case at $\bar{I} = 0.30$ bits, using values from Table VI and marginal $q^* = 0.163$:

$$\Delta I_1 = 0.10 [0.982 \ln(0.982/0.163) + 0.018 \ln(0.018/0.837)]$$

$$= 0.10 \times 1.694 \text{ nats} = 0.244 \text{ bits},$$

$$\Delta I_2 = 0.7075 [0.085 \ln(0.085/0.163) + 0.915 \ln(0.915/0.837)]$$

$$= 0.027 \text{ bits},$$

$$\Delta I_3 = 0.1925 [0.024 \ln(0.024/0.163) + 0.976 \ln(0.976/0.837)]$$

$$= 0.029 \text{ bits},$$

summing to $0.244 + 0.027 + 0.029 = 0.300$ bits, confirming Table III in the main text.

B.7. Saturation budget and the declining ratio

The frontier $V(\bar{I})$ rises until the full-information value $U_\infty - U(0)$ is reached and is flat beyond. The smallest capacity at which this happens is the saturation budget $\bar{I}_{\text{sat}}$. Since the optimal action under full information, $\alpha^\circ(\theta)$, is unique in every state, any channel that attains the full-information value must play $\alpha^\circ(\Theta)$ with probability one, and the cheapest such channel is the deterministic rule itself. Hence

$$\bar{I}_{\text{sat}} = I(\Theta; \alpha^\circ(\Theta)) = H(\alpha^\circ(\Theta)),$$

the entropy of the full-information action, which is also the limit of (34) as $\lambda \rightarrow 0$.

- **Symmetric:** the investor exits in θ_1 and invests in θ_2 , each with probability $\frac{1}{2}$, so

$$\bar{I}_{\text{sat}}^{\text{sym}} = 1 \text{ bit.}$$

- **Skewed:** the investor exits only in the crash state, which has probability 0.10. As $\lambda \rightarrow 0$

the Gibbs signal perfectly separates the crash state from the others and $q^* \rightarrow \mu_1 = 0.10$,

$$\text{so } \bar{I}_{\text{sat}}^{\text{sk}} = H(\mu_1) = -0.10 \log_2(0.10) - 0.90 \log_2(0.90) \approx 0.469 \text{ bits.}$$

The lower saturation budget of the skewed problem does not reflect lower uncertainty. The symmetric prior has entropy $\ln 2 = 0.693$ nats = 1 bit, whereas the skewed prior $(0.10, 0.7075, 0.1925)$ has

$$H^{\text{sk}} = - \sum_i \mu_i \ln \mu_i = 0.792 \text{ nats} = 1.143 \text{ bits},$$

which is strictly larger. Nor does it depend on the CARA weighting of the Gibbs kernel: it is fixed by the distribution of the full-information action. Telling the normal state from the

gain state has no value, because the investor stays invested in both, so only the information needed to detect the crash is worth acquiring.

This explains the declining ratio in Table II. At $\bar{I} = 0.10$ bits, the skewed problem uses $0.10/0.469 = 21\%$ of its saturation budget and crash-detection signals remain highly valuable; the ratio is 2.24. At $\bar{I} = 0.30$ bits it uses $0.30/0.469 = 64\%$ and marginal information value falls rapidly; the ratio declines to 1.85. At $\bar{I} = 0.469$ bits the skewed VoI reaches its full-information value $U_{\infty}^{\text{sk}} - U^{\text{sk}}(0) = 0.1178$, and the ratio is about 1.58. Beyond that only the symmetric VoI rises, and the ratio falls to the full-information ratio $0.1178/0.1064 = 1.107$ at $\bar{I} = 1$ bit, where both problems are saturated.

B.8. Nonlinear scaling with risk aversion: parameter sensitivity

Sensitivity to γ . The coefficient $\gamma = 2$ is a conventional benchmark in portfolio and decision-theory applications. Since CARA utility $u_i(\gamma) = -e^{-\gamma(r_f + R_i)}$ is nonlinear in γ , the VoI level varies with γ for both distributions. Table IV of the main text therefore reports results for $\gamma \in \{1, 2, 4\}$ to confirm that the qualitative finding, that skewed distributions command a substantial information premium over symmetric benchmarks, is stable across this range rather than an artifact of any particular choice. The ratio between skewed and symmetric VoI also varies with γ , so no single number characterises the premium; the table reports the full range.

Which parameters are critical. The Gibbs concentration on the crash state is governed by the exponent $(u_0 - u_1^{\text{sk}})/\lambda^*$ (equation (41)). Decomposing this:

$$\frac{u_0 - u_1^{\text{sk}}}{\lambda^*} = \frac{e^{-\gamma r_f} (e^{-\gamma R_1} - 1)}{\lambda^*}, \quad (46)$$

where λ^* is itself determined by the target $\bar{I}$. Three parameters drive the magnitude of the premium.

First, the crash excess return $R_1 = -40\%$ is the single most important parameter: it enters the exponent via $e^{-\gamma R_1} - 1$, which grows rapidly in $|\gamma R_1|$, so larger crashes and higher risk aversion both amplify concentration on the crash state. R_1 and μ_1 are chosen to represent a severe crash year; together with the other two states they imply a skewness of -1.54 , more pronounced than in annual S&P 500 returns, so the example is illustrative rather than calibrated. A smaller crash magnitude ($|R_1|$ closer to zero) would reduce concentration

and compress the ratio; a larger crash (consistent with longer-horizon annual data) would amplify it.

Second, the crash probability $\mu_1 = 10\%$ determines how much of the concentrated signal capacity actually enters total mutual information. Formally, the crash state contributes ΔI_1 as defined in (45) to the total budget $\bar{I}$, that is, the crash probability scales the information content of the crash-state signal directly. Reducing μ_1 compresses ΔI_1 mechanically even when the per-state Gibbs weight is extreme; for the same reason the saturation budget of the skewed problem, $H(\mu_1) = 0.469$ bits, is below the symmetric 1 bit.

Third, the information budget $\bar{I}$ relative to the skewed saturation budget of 0.469 bits determines where the agent sits on the diminishing-returns curve: at $\bar{I} = 0.10$ bits (21% of the skewed saturation budget) the ratio is 2.24; at $\bar{I} = 0.30$ bits (64%) it has fallen to 1.85. The level of γ , by contrast, shifts the VoI level strongly for both distributions but has a more modest effect on the *ratio*, as Table IV confirms.

B.9. Summary

All computations follow from the utility values in (4)–(9), the self-consistent marginals in (27) and (32), the mutual information formula (34), and the value function via the primal–dual identity (22), with the shadow price λ^* located by bisecting the monotone residual (33). The entire calculation involves only exponentiation, logarithms, and polynomial root-finding on a state space of size at most three.

APPENDIX: REFERENCES

- Ahmadi-Javid, A. (2012): “Entropic Value-at-Risk: A New Coherent Risk Measure,” *Journal of Optimization Theory and Applications*, 155, 1105–1123. [2, 5]
- Arrow, K. J. (1996): “The Economics of Information: An Exposition,” *Empirica*, 23, 119–128. [6]
- Artzner, P., F. Delbaen, J.-M. Eber, and D. Heath (1999): “Coherent Measures of Risk,” *Mathematical Finance*, 9, 203–228. [4]
- Azrieli, Y. (2021): “Constrained versus Unconstrained Rational Inattention,” *Games*, 12, 3. [4]
- Behringer, S. (2026): “Hard Constraints, Convex Duality, and the Endogenous Cost of Information,” Working Paper. [1, 2, 3, 4, 5, 6, 10, 11, 15]
- Behringer, S., and R. V. Belavkin (2026): *The Value of Information in Economics and Decision Theory: A Continuous Framework from Shannon to Stratonovich*. Book manuscript. [7]
- Cover, T. M., and J. A. Thomas (2006): *Elements of Information Theory* (Second Ed.). Hoboken: Wiley. [11]

- De Oliveira, H., T. Denti, M. Mihm, and K. Ozbek (2017): “Rationally Inattentive Preferences and Hidden Information Costs,” *Theoretical Economics*, 12, 621–654. [4]
- Denti, T., M. Marinacci, and L. Montrucchio (2020): “A Note on Rational Inattention and Rate Distortion Theory,” *Decisions in Economics and Finance*, 43, 75–89. [2, 3]
- Donsker, M. D., and S. R. S. Varadhan (1975): “Asymptotic Evaluation of Certain Markov Process Expectations for Large Time, I,” *Communications on Pure and Applied Mathematics*, 28, 1–47. [2, 11]
- Föllmer, H., and A. Schied (2002): “Convex Measures of Risk and Trading Constraints,” *Finance and Stochastics*, 6, 429–447. [4]
- Föllmer, H., and A. Schied (2016): *Stochastic Finance: An Introduction in Discrete Time* (Fourth Ed.). Berlin: De Gruyter. [2, 4]
- Frittelli, M., and E. Rosazza Gianin (2002): “Putting Order in Risk Measures,” *Journal of Banking and Finance*, 26, 1473–1486. [4]
- Grechuk, B., A. Malandii, R. T. Rockafellar, and S. Uryasev (2026): “The Risk Quadrangle in Optimization: An Overview with Recent Results and Extensions,” arXiv:2603.27370. [5]
- Hansen, L. P., and T. J. Sargent (2001): “Robust Control and Model Uncertainty,” *American Economic Review*, 91, 60–66. [4]
- Heal, G., and A. Millner (2014): “Uncertainty and Decision Making in Climate Change Economics,” *Review of Environmental Economics and Policy*, 8, 120–137. [9]
- Matějka, F., and A. McKay (2015): “Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model,” *American Economic Review*, 105, 272–298. [1, 2, 3]
- Pindyck, R. S. (2013): “Climate Change Policy: What Do the Models Tell Us?” *Journal of Economic Literature*, 51, 860–872. [10]
- Sims, C. A. (2003): “Implications of Rational Inattention,” *Journal of Monetary Economics*, 50, 665–690. [1]
- Stratonovich, R. L. (2020): *Theory of Information and its Value*, ed. by R. V. Belavkin, P. M. Pardalos, and J. C. Principe. Cham: Springer. Russian original 1975. [1, 6, 15]