

Hard Constraints, Convex Duality, and the Endogenous Cost of Information

Stefan Behringer¹

Value of Information Group, Universität Bielefeld, Germany

Abstract

Rational inattention (RI) has two formulations: a hard constraint bounding mutual information by a Shannon capacity, and a soft constraint penalising it linearly. We relate them by convex duality. First, the soft-constraint objective is the Fenchel conjugate of the Stratonovich/Shannon Value of Information (VoI), the gain in expected utility from a capacity; the VoI is recovered from it, and the marginal cost of information at a given capacity is its slope. These results hold for a broad class of costs, including f -divergence generalisations of mutual information, and a single linear penalty represents the hard constraint across budgets only where the VoI is affine. Second, for Shannon costs a hard-constrained agent chooses as an RI logit agent whose price of information is the inverse of its own demand, and the VoI is the minimum over prices of the expected logsum plus the cost of capacity. Third, scaling the stakes scales the frontier, which yields each formulation's

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

¹The author is grateful to Roman V. Belavkin, Julien Béguinot, 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 reviewers for comments that motivated some results of this paper.

information response: none under a fixed capacity, the full demand response under a linear penalty, and an intermediate one under a convex cost of capacity, whose marginal cost the response identifies. With an endowment of capacity and a market for more, information responds only above a threshold stake, which reveals the shadow price of the hard constraint. The Gaussian quadratic benchmark displays the results in closed form.

Keywords: rational inattention, Value of Information, hard and soft information constraints, convex duality, Fenchel conjugate, rate distortion theory, stakes, Gaussian quadratic benchmark

JEL classification: D81, D83, C61

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.

For finite Shannon problems, [Azrieli \(2021\)](#) shows that a policy solves the constrained problem if and only if the constraint binds and the policy solves the penalised problem for some $\lambda > 0$ (his Proposition 1); that the multiplier $\lambda(c)$ is unique for each capacity c , continuous and weakly decreas-

ing (his Lemma 1), and that it may be constant on an interval of capacities, which corresponds to ‘regime changes’ in the penalised problem (his Section 3 and Propositions 2 and 3); that the two versions have opposite comparative statics, changes in the prior generically moving the optimal posteriors of the constrained problem but, locally, not those of the penalised one, and changes in the stakes moving the penalised solution but not the constrained one (his Propositions 5 and 6); and that, unless the prior is an indifference point, the penalised problem selects no information above a finite price (his Proposition 7). [Matějka and McKay \(2015, p. 282\)](#) make the same point about stakes: under the linear penalty the agent can pay more attention when the stakes are high, and [Maćkowiak, Matějka, and Wiederholt \(2023, Section 2.4.2\)](#) note that in a static setup the two problems are dual, so that the solution under a fixed capacity is again a logit, with λ the endogenous shadow cost of information. These results compare optimal policies through the Kuhn–Tucker conditions. At the level of value functions the two formulations are related by the conjugate duality of [Rockafellar \(1974\)](#), which [Denti, Marinacci, and Montruccio \(2020\)](#) apply to finite Shannon problems. What has not been worked out is what this duality implies: which conclusions survive without the linear entropy penalty, for which class of information costs, and how the two formulations compare across decision problems, where each must hold a different object fixed.

This paper addresses these questions and makes three contributions.

The first contribution is to determine the scope of this duality. The Stratonovich/Shannon Value of Information (VoI) $V(\bar{I})$, the gain in expected utility attainable with an information capacity $\bar{I}$, is the information-utility

frontier of rate distortion theory in the sense of [Stratonovich \(1975/2020\)](#), and the soft-constraint objective is its Fenchel conjugate (Theorem 1 and (6)), an instance of [Rockafellar \(1974, Theorem 7\)](#) that holds for any cost functional. The VoI is recovered from the soft-constraint objective by biconjugation (Theorem 2), and the marginal cost of information at a given capacity is the slope of V (Theorem 3), determined by the utility function, the prior and the action space. Both follow from Theorem 1 by standard convex analysis once V is concave and continuous; what the paper supplies is the class of costs for which this holds (Lemma 4). It consists of every information cost that, like mutual information, never rises when a signal is garbled, is convex in the signal structure and is lower semicontinuous, and it includes the f -divergence analogues of mutual information and the posterior-separable costs of [Caplin, Dean, and Leahy \(2022\)](#) whose generalised entropy is continuous. The shadow price of capacity is independent of the budget only on intervals where V is affine, and on such intervals the soft formulation does not determine the information acquired (Proposition 2). This recasts the constant-multiplier intervals of [Azrieli \(2021, Propositions 2 and 3\)](#) as the affine pieces of the frontier and extends them beyond Shannon costs.

The second contribution is to connect the duality to the RI logit of [Matějka and McKay \(2015\)](#), the solution of the soft problem for Shannon costs. Its value, the expected logsum (inclusive value), is the Fenchel conjugate of the VoI frontier, and the information it acquires at each price defines a downward-sloping demand for information whose inverse is the shadow price of the hard constraint. That a hard-constrained agent chooses as an RI logit agent at an endogenous price is noted by [Maćkowiak, Matějka, and Wieder-](#)

[holt \(2023, Section 2.4.2\)](#); the duality identifies this price as the inverse of the agent’s own demand for information, and the frontier as the minimum over prices of the expected logsum plus the cost of capacity (Proposition 3). Because the price depends on the decision problem, a logit model with a single price fitted across decision problems that share a capacity is misspecified unless their shadow prices at that capacity happen to coincide.

The third contribution takes up the comparison across stakes. Neither formulation is free of primitives: the hard constraint fixes the capacity, the soft constraint the price or cost schedule, and across decision problems they predict different behaviour. Scaling the stakes of a decision problem scales its frontier, so the frontier of the unscaled problem delivers the information response of every formulation (Proposition 4): none under a fixed capacity, as in [Azrieli \(2021, Proposition 6\)](#), a fall of the effective price from λ to λ/k under a linear penalty, and a monotone response under a convex cost of capacity, which lies between the two: calibrated to the same choice at one stake, it moves capacity by less than the linear penalty in either direction. Once capacity can be traded, so that the agent holds an endowment and can buy more at a price, the information acquired does not respond to the stakes below a threshold and responds as under a linear penalty above it, so that the two formulations are complementary regimes of one model, and the threshold stake reveals the shadow price of the endowment (Corollary 1). The response of information to the stakes also identifies the marginal cost of capacity at every capacity chosen as the stakes vary; in the experiments of [Dean and Neligh \(2023\)](#) it lies between the polar cases, as under a convex cost of capacity. [Sims’s \(2006\)](#) distinction between processing capacity,

the ‘wiring’ that brings information in and costly investigation in financial markets illustrates both regimes, and automated trading makes the capacity constraint literal.

The Gaussian quadratic benchmark, called the linear-quadratic case by [Sims \(2006\)](#), displays these results in closed form (Section 4), and by [Shannon’s \(1948\)](#) bound the Gaussian frontier never overstates the VoI of a prior with the same variance.

The VoI also generates a class of convex risk measures, and the framework has applications to diagnostic testing, portfolio choice under tail risk, and climate policy under deep uncertainty. These are developed in the companion paper [Behringer \(2026a\)](#), so that the present paper can concentrate on the duality structure itself.

The paper connects several literatures that have developed largely independently. RI 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, and [Denti, Marinacci, and Montrucchio \(2020\)](#) translate tools from rate distortion theory into the language of RI. For Shannon costs on general state and action spaces, they show that the soft problem is equivalent to a maximisation over the marginal distribution of actions (their Theorem 1), and for finite problems they derive the rate distortion duality from the saddle property of the Lagrangian (their Proposition 11). Their note does not go beyond Shannon costs, prove the concavity and continuity of the frontier, or characterise the multipliers as its subgradients, which is what the present paper adds. The VoI, rooted in classical rate distortion theory

(Stratonovich, 1975/2020), is developed for economic contexts in Behringer and Belavkin (2023, 2025, 2026a). Robust control approaches (Hansen and Sargent, 2001, 2008) share the entropy-based structure but optimise over alternative probability models rather than over signal distributions. De Lara and Gossner (2020) exploit the duality between actions and beliefs to characterise the value function via convex analysis and derive conditions under which information has positive value 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. A hard bound on Shannon mutual information, itself a posterior-separable functional, does not escape this critique. The duality results of Section 3, however, do not depend on Shannon costs and apply to any experiment-based cost in the class described above.

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) in-

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 to impose or recover, taking it as a primitive of the model. For any information cost in the class of Section 3, Theorems 2 and 3 map one formulation into the other problem by problem, and Section 6 shows how the frontier disciplines the comparison across problems. A companion paper, Behringer (2026b), studies costs that are increasing convex transforms of an information measure and the entropic certainty equivalent of the soft-constraint value.

Section 2 defines the two formulations and develops the framework. Section 3 states the duality results and their scope. Section 4 treats the Gaussian quadratic benchmark. Section 5 relates hard-constrained choice to the RI logit. Section 6 compares the formulations across decision problems that differ in their stakes. Section 7 concludes. Appendix A collects the proofs, and Appendix B derives the duality results from the conjugate duality of Rockafellar (1974).

2. Framework

2.1. Basic Setup

Let Ω be a finite state space and $\mathcal{A}$ a finite action space. The state of the world is a random variable Θ with prior distribution $\mu \in \Delta(\Omega)$, whose realisations are denoted θ ; throughout, upper-case letters denote random variables and lower-case letters their realisations, and $E\{\cdot\}$ is the expectation

under the joint distribution of all random variables involved. The agent chooses a signal structure, a conditional distribution $p(s|\theta)$ of the signal S given the state (a *channel* in the language of information theory), and after observing the realisation s selects the Bayesian optimal action

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

where $U : \mathcal{A} \times \Omega \rightarrow \mathbb{R}$ is the utility function. For a fixed prior μ , Shannon mutual information between Θ and S depends only on the signal structure, and we write it as a function of p :

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

measured in bits (or in nats when $\log = \ln$), where $p(\theta, s) := \mu(\theta)p(s|\theta)$ is the joint distribution, $p(s) := \int \mu(\theta)p(s|\theta) d\theta$ is the marginal signal distribution, and integrals over θ are sums over the finite state space. Section 4 departs from finiteness and works with a real-valued state and real-valued actions.

2.2. The Stratonovich/Shannon Value of Information

The central object of this paper is built from the supremum of expected utility subject to a mutual information constraint:

$$U(\bar{I}) := \sup_{p(s|\theta)} [E\{U(a^*(S), \Theta)\} \text{ subject to } I(p) \leq \bar{I}]. \quad (2)$$

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

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

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

2.3. The Two Formulations and Local Equivalence

Definition 1 (Hard and Soft Information Constraints).

1. *Hard constraint (Shannon capacity):*

$$U(\bar{I}) = \max_{p(s|\theta)} E\{U(a^*(S), \Theta)\} \quad \text{subject to} \quad I(p) \leq \bar{I}, \quad (4)$$

where $\bar{I} > 0$ is an exogenous information capacity bound. This is the value (2), with the supremum attained (Lemma 3).

2. *Soft constraint (entropy penalty):*

$$V_{\text{soft}}(\lambda) := \max_{p(s|\theta)} \{E\{U(a^*(S), \Theta)\} - \lambda I(p)\}, \quad (5)$$

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 $R_+(I) \leq R_0$ so that $V(I) \geq 0$. Our formulation (2) is the utility-maximisation counterpart. Throughout this paper we use λ for the Lagrange multiplier on the capacity constraint, following the conventions of the RI literature. In the logit choice probabilities (10) λ plays the role of a temperature, $\lambda = k_B T$, since it divides the payoff in the exponent. [Stratonovich \(1975/2020, Section 9.4\)](#) instead parametrises the extremal distribution by $\beta = 1/\lambda$, the inverse temperature of statistical mechanics, writing it as proportional to $\exp(-\beta c)$ for cost $c = -U$.

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

The Lagrangian of (4) is $\mathcal{L} = E\{U(a^*(S), \Theta)\} - \lambda[I(p) - \bar{I}]$, which is structurally identical to (5) for the same λ . The first-order conditions therefore coincide for any fixed λ . By KKT complementary slackness, $\lambda \cdot [I(p) - \bar{I}] = 0$ with $\lambda > 0$ implies the constraint binds at the optimum.

The multiplier is a slope of the VoI. Since V is concave (Lemma 4 in Appendix B), we call λ a *subgradient* of V at $\bar{I}$ if $V(J) \leq V(\bar{I}) + \lambda(J - \bar{I})$ for all $J \geq 0$, and write $\partial V(\bar{I})$ for the set of subgradients, the *subdifferential*; where V is differentiable, $\partial V(\bar{I}) = \{V'(\bar{I})\}$. For a convex function such as V_{soft} below, subgradients and the subdifferential are defined with the inequality reversed, as usual.

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

Proposition 1 is the Kuhn–Tucker connection of Azrieli (2021, Proposition 1), stated with subgradients so that it covers budgets at which V is not differentiable.

Remark. *Local equivalence concerns a fixed budget. For a convex, increasing cost schedule $c(I)$ applied to mutual information, that is, a cost functional $C(p) = c(I(p))$ of the kind considered in Theorem 1, the soft problem with cost $C(p)$ selects the budget $\bar{I}$ at which $c'(\bar{I}) \in \partial V(\bar{I})$ and, at that budget, has the same solution as the hard problem; the schedule c , however, has to be specified. This reduction is also Proposition 1 of Behringer (2026b). Costs*

that are non-linear transforms of mutual information are used by [Paciello and Wiederholt \(2014, p. 363\)](#), who measure attention by mutual information and assume its cost increasing and convex, with a linear and an exponential cost as examples (p. 372), and discussed by [Dean and Neligh \(2023, Section V\)](#), and [Maćkowiak, Matějka, and Wiederholt \(2023, Section 2.4.2\)](#) note that fatigue or decreasing returns to time spent on other activities can make the appropriate cost convex in entropy. The hard-constraint formulation needs no schedule but fixes the capacity instead; [Section 6](#) compares the two across stakes.

2.4. Properties of the VoI

The VoI $V(\bar{I})$ is non-decreasing, concave and continuous on $[0, \infty)$ ([Lemma 4](#) in [Appendix B](#)), reflecting diminishing returns to information: the first bit of information is at least as valuable as subsequent bits. Its subdifferential $\partial V(\bar{I})$ is non-empty at every $\bar{I} > 0$.

Information also has a finite saturation point. Since $U(\bar{I}) \leq E\{\max_a U(a, \Theta)\}$ for every budget, with equality once the agent can afford to learn a state-optimal action $a^o(\theta) \in \arg \max_a U(a, \theta)$ in each state, the *saturation budget*

$$\bar{I}^{\text{sat}} := \min\{\bar{I} \geq 0 : U(\bar{I}) = E\{\max_a U(a, \Theta)\}\}$$

is determined by U , μ and $\mathcal{A}$ alone. It is finite: the channel that recommends $a^o(\Theta)$ attains the full-information value, and its mutual information is $H(a^o(\Theta))$, so that

$$\bar{I}^{\text{sat}} \leq H(a^o(\Theta)) \leq \min\{H(\Theta), \log |\mathcal{A}|\},$$

where the second inequality holds because $a^o(\Theta)$ is a function of Θ taking at most $|\mathcal{A}|$ values. The minimum in the definition exists because U is

continuous, so that the budgets attaining the full-information value form a closed set. By Lemma 4, U is strictly increasing on $[0, \bar{I}^{\text{sat}}]$ and constant on $[\bar{I}^{\text{sat}}, \infty)$: if $U(\bar{I}_1) = U(\bar{I}_2)$ for some $\bar{I}_1 < \bar{I}_2$, concavity and monotonicity make U constant on $[\bar{I}_1, \infty)$, so that $U(\bar{I}_1)$ is already the full-information value and $\bar{I}_1 \geq \bar{I}^{\text{sat}}$. Hence for $0 < \bar{I} < \bar{I}^{\text{sat}}$ the capacity constraint binds and every $\lambda^* \in \partial V(\bar{I})$ is strictly positive, while for $\bar{I} > \bar{I}^{\text{sat}}$ the constraint is slack and $\partial V(\bar{I}) = \{0\}$. In the continuous-state Gaussian benchmark of Section 4, full information cannot be reached with finite mutual information, and we set $\bar{I}^{\text{sat}} = \infty$.

3. Convex Duality and Endogenous Information Costs

3.1. The Fenchel Conjugate Relationship

Theorem 1 (Fenchel–Legendre Representation). *For any cost functional $0 \leq C(p) < \infty$ of the signal structure $p(s|\theta)$, define*

$$U_C(\bar{I}) := \sup_{p(s|\theta)} \{E\{U(a^*(S), \Theta)\} : C(p) \leq \bar{I}\},$$

$$V_{\text{soft}, C}(\lambda) := \sup_{p(s|\theta)} \{E\{U(a^*(S), \Theta)\} - \lambda C(p)\},$$

with $\sup \emptyset = -\infty$. Then for every $\lambda \geq 0$,

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

For Shannon costs, $C = I$, the functions U_C and $V_{\text{soft}, C}$ are U and V_{soft} , and since $U(0)$ does not depend on $\bar{I}$, Theorem 1 reads

$$V_{\text{soft}}(\lambda) = \sup_{\bar{I} \geq 0} \{U(\bar{I}) - \lambda \bar{I}\} = \sup_{\bar{I} \geq 0} \{V(\bar{I}) - \lambda \bar{I}\} + U(0). \quad (6)$$

Up to the constant $U(0)$, the soft-constraint objective is therefore the Fenchel conjugate of the VoI: $V_{\text{soft}}(\lambda) - U(0) = (-V)^*(-\lambda)$, the convex conjugate of $-V$ evaluated at $-\lambda$.

Remark (Scope of the duality results). *The proof of Theorem 1 uses nothing about C beyond $C(p) \in [0, \infty)$: it only interchanges the supremum over signal structures with the supremum over budgets. The theorem is also an instance of the conjugate duality of Rockafellar (1974, Theorem 7), with perturbation function $F(p, \bar{I}) = -E\{U(a^*(S), \Theta)\} + \delta[C(p) \leq \bar{I}]$, where $\delta[\cdot]$ is the $\{0, \infty\}$ -valued indicator; Appendix B carries out this derivation step by step. We are grateful to a referee for this observation. Theorems 2 and 3 additionally require the VoI to be concave and continuous (Lemma 4). This follows from three properties of mutual information: garbling a signal never raises its cost (the data-processing inequality), a mixture of two signal structures costs at most the average of their costs (convexity in the channel for a fixed prior), and cost varies continuously with the signal structure (lower semi-continuity suffices). Any information cost with these properties, such as the f -divergence analogues of mutual information, supports the same argument, so the results of this section apply to that entire class, not only to Shannon costs. It contains the posterior-separable costs of Caplin, Dean, and Leahy (2022), among them the neighbourhood-based costs of Hébert and Woodford (2021), whenever their generalised entropy is continuous on the simplex; the log-likelihood ratio cost of Pomatto, Strack, and Tamuz (2023) also has the three properties but is infinite for channels under which some signal rules*

out a state.³ Mutual information is specific to what follows: the closed-form Gaussian benchmark (Section 4) and the link to the RI logit (Proposition 3) rely on the logarithmic structure of Shannon information, which is also what connects the framework to Sims (2003) and the RI literature.

3.2. Biconjugate Recovery

Theorem 2 (Biconjugate Recovery). *For every $\bar{I} \geq 0$,*

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

equivalently

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

By Lemma 4, the VoI is concave and continuous on $[0, \infty)$, so $-V$ is closed and convex. Theorem 2 is the biconjugate identity applied to $-V$ (Rockafellar, 1974, Theorem 5, p. 16): $(-V)^{**} = \text{cl co}(-V) = -V$, where cof denotes the convex envelope and cl its closure, both of which leave $-V$ unchanged;⁴ by (6), the conjugate of $-V$ is the soft-constraint objective less the constant

³For both families convexity in the channel follows from monotonicity under garbling. A mixture of two channels on a common signal space is a garbling of the labelled mixture, in which the agent also observes which channel was used, and the labelled mixture costs exactly the average of the two costs.

⁴The convex envelope cof is the largest convex function below f , and the closure $\text{cl}f$ the largest lower semicontinuous function below f , so each returns f when f already has the property. Here $-V$, extended by $+\infty$ for $\bar{I} < 0$, is convex because V is concave, and lower semicontinuous because V is continuous on $[0, \infty)$, including from the right at $\bar{I} = 0$ (Lemma 4).

$U(0)$. No functional form is assumed for the cost of information as a function of $I(p)$: the implied cost schedule is recovered from the VoI rather than imposed.

3.3. Endogenous Identification of the Marginal Cost

Theorem 3 (Endogenous Marginal Cost). *At any $\bar{I} \geq 0$,*

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

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, Corollary 12A, p. 35). It states that λ^* is a subgradient of V at $\bar{I}$ if and only if $-\bar{I}$ is a subgradient of V_{soft} at λ^* : 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). *At a given budget, the shadow price of capacity $\lambda^*(\bar{I}) \in \partial V(\bar{I})$ is determined by U , μ , and $\mathcal{A}$. It is independent of the budget on an interval of budgets only if V is affine on that interval. A single exogenously imposed λ therefore represents the hard constraint across an interval of budgets only in the special case in which V is affine there. Where V is affine with slope λ on an interval, the soft problem at that λ is solved at every budget in the interval, so that the soft formulation does not determine the information acquired.*

Proposition 2 recasts the constant-multiplier intervals of Azrieli (2021, Propositions 2 and 3) at the level of the frontier: they are exactly the affine

pieces of V and the kinks of V_{soft} . It also gives a formal reading of Sims’s (2003; 2006) advocacy of the hard-constraint approach: a price is tied to a budget, so a linear penalty cannot stand in for a capacity across budgets. This does not favour the hard formulation across decision problems, where it must hold the capacity fixed just as the soft formulation holds the price fixed (Section 6).

Sims (2006, Section C, p. 160) distinguishes three kinds of information cost. Human information processing and the ‘wiring’ that brings information in, such as subscriptions and data lines, are reasonably measured in bits, with costs linear or at least smooth in bits but at quite different prices per bit; the price of processing capacity is its shadow price in the agent’s overall problem. Costly investigation, by experimentation, surveying or exploration, is different: a single bit may be very cheap or very expensive to obtain, so its cost is probably not well measured in bits, and in asset markets the uncertainty of sophisticated investors is dominated by information that is not freely available. A linear penalty $\lambda I(p)$ is thus appropriate at best for processing or wiring, each at its own price. For processing capacity the hard formulation supplies that price as the slope of the VoI, derived from the decision problem and the capacity with no functional form assumed for the cost of information. For costly investigation the results of this section apply to a cost functional C other than mutual information, provided it has the three properties listed after Theorem 1.

4. The Gaussian Quadratic Benchmark

The Gaussian quadratic case, called the linear-quadratic case by [Sims \(2006\)](#),⁵ is the canonical example of rate distortion theory because it is one of the rare cases admitting a closed-form frontier ([Shannon, 1948](#); [Sims, 2003](#); [Cover and Thomas, 2006](#)). This section uses it to display the duality results of Section 3 in closed form, and clarifies what distinguishes the Gaussian prior from non-Gaussian priors with the same variance.

Throughout this section the state Θ is real-valued with prior mean μ_0 and finite variance σ^2 , actions are real numbers, utility is quadratic, $U(a, \theta) = -(a - \theta)^2$, and information is measured in nats. Theorem 1 does not use finiteness of the state space, and the Gaussian frontier below is explicit, concave and continuous, so Theorems 2 and 3 apply to it directly.

4.1. Quadratic Utility and the Distortion-Rate Function

Define the distortion-rate function of the prior under squared error,

$$D(\bar{I}) := \inf_{q(a|\theta)} \{E\{(A - \Theta)^2\} : I(\Theta, A) \leq \bar{I}\}, \quad (7)$$

where A is a reconstruction of the state drawn from the channel $q(a | \theta)$.

Lemma 1 (Quadratic Utility). *Let the prior have $\sigma^2 < \infty$ and let $U(a, \theta) = -(a - \theta)^2$.*

- (i) *For every signal S , $E\{U(a^*(S), \Theta)\} - U(0) = \sigma^2 - E\{\text{Var}(\Theta | S)\}$.*

⁵In the Gaussian quadratic setting the optimal signal is a linear function of the state, and the optimal action is a linear function of the signal, since the posterior mean is linear in the observation under Gaussian conjugacy.

(ii) For every $\bar{I} \geq 0$, $U(\bar{I}) = -D(\bar{I})$ and $V(\bar{I}) = \sigma^2 - D(\bar{I})$.

Part (i) holds for every prior and every signal structure: under quadratic utility, the gain from a signal is the reduction in expected posterior variance whatever the distribution of the state. Dependence of the VoI on variance reduction alone is therefore not a property that singles out the Gaussian case. What distinguishes priors is how much expected posterior variance a given number of nats can remove. Part (ii) shows that this is measured by the distortion-rate function D , so that V is the VoI of [Stratonovich \(1975/2020, eq. \(9.3.7\)\)](#) for squared-error cost and, by [Theorem 3](#), the shadow price of capacity is the marginal variance reduction, $\lambda^*(\bar{I}) = -D'(\bar{I})$ where D is differentiable.

4.2. The Gaussian Prior: an Explicit Conjugate Pair

Lemma 2 (Gaussian Frontier and its Conjugate). *Let $\Theta \sim N(\mu_0, \sigma^2)$ and $U(a, \theta) = -(a - \theta)^2$.*

(i) For every $\bar{I} \geq 0$,

$$V(\bar{I}) = \sigma^2(1 - e^{-2\bar{I}}), \quad (8)$$

attained by the Gaussian signal $S = \Theta + \varepsilon$ with $\varepsilon \sim N(0, \sigma_\varepsilon^2)$ independent of Θ and $\sigma_\varepsilon^2 = \sigma^2/(e^{2\bar{I}} - 1)$ for $\bar{I} > 0$. The frontier is strictly increasing and strictly concave, $\bar{I}^{\text{sat}} = \infty$, and the shadow price of capacity is $\lambda^*(\bar{I}) = V'(\bar{I}) = 2\sigma^2 e^{-2\bar{I}}$.

(ii) The soft-constraint objective is

$$V_{\text{soft}}(\lambda) - U(0) = \begin{cases} \sigma^2 - \frac{\lambda}{2} - \frac{\lambda}{2} \ln \frac{2\sigma^2}{\lambda}, & 0 < \lambda \leq 2\sigma^2, \\ 0, & \lambda \geq 2\sigma^2, \end{cases} \quad (9)$$

and the capacity chosen at price λ is $I(\lambda) = \frac{1}{2} \ln(2\sigma^2/\lambda)$ for $\lambda \leq 2\sigma^2$ and $I(\lambda) = 0$ otherwise.

Three features of Lemma 2 stand out. First, $I(\lambda)$ is the Gaussian counterpart of the demand for information of Proposition 3(i), and the conjugate pair identifies budgets and prices mutually, as in Theorem 3: differentiating (9) gives $V'_{\text{soft}}(\lambda) = -I(\lambda)$, and $I(\lambda)$ and $\lambda^*(\bar{I})$ are inverse functions. Second, the price $2\sigma^2 = V'(0)$ is a choke price: it is the marginal value of the first nat, and an agent facing a soft penalty $\lambda \geq 2\sigma^2$ acquires no information, the Gaussian counterpart of the finite choke price of Azrieli (2021, Proposition 7). Third, the shadow price falls by the factor e^{-2} with every additional nat of capacity and scales with prior uncertainty σ^2 . A penalty λ calibrated to one budget $\bar{I}_0$ therefore misprices capacity at any other budget $\bar{I}_1$ by the factor $e^{-2(\bar{I}_1 - \bar{I}_0)}$, and a penalty calibrated to one decision problem misprices it, at the same budget, in any problem with a different prior variance. This is Proposition 2 in its simplest explicit form.

Remark (The Gaussian frontier as a lower bound). *The signal of Lemma 2 achieves the Gaussian variance reduction for any prior with the same variance (Appendix A), which is Shannon’s (1948) upper bound on the distortion-rate function of a source of given variance. Hence, under quadratic utility,*

$$V(\bar{I}) \geq \sigma^2(1 - e^{-2\bar{I}}) \quad \text{for all } \bar{I} \geq 0$$

for every prior with variance σ^2 : the Gaussian benchmark never overstates the VoI. A two-point prior on $\mu_0 \pm \sigma$ with equal probabilities illustrates the size of the gap: it is fully revealed with $\ln 2$ nats, so that $V(\ln 2) = \sigma^2$, whereas the Gaussian frontier gives $\frac{3}{4}\sigma^2$ at the same budget. Calibrating the Gaussian

benchmark to the variance of a non-Gaussian state therefore understates the VoI, by an amount that depends on the distortion-rate function of the actual prior. For a prior with a density, [Shannon's](#) entropy-power bound $D(\bar{I}) \geq Ne^{-2\bar{I}}$, where $N = e^{2h(\Theta)}/(2\pi e) \leq \sigma^2$ is the entropy power of the prior, bounds the amount:

$$\sigma^2(1 - e^{-2\bar{I}}) \leq V(\bar{I}) \leq \sigma^2 - Ne^{-2\bar{I}},$$

with equality throughout for a Gaussian prior, for which $N = \sigma^2$. [Belavkin, Pardalos, and Principe \(2024, eqs. \(7\)–\(8\)\)](#) give the frontier for squared error in terms of the entropy of the state and, in its Gaussian form calibrated to the variance, use it as a benchmark for forecasts of heavy-tailed cryptocurrency returns. For such a state $N < \sigma^2$, and by the first inequality the Gaussian benchmark overstates the smallest mean squared error attainable with a given amount of information. The consequences for tail risk are analysed in [Behringer \(2026a\)](#).

5. Hard-Constrained Choice: Logit at the Shadow Price

For Shannon costs the soft problem has a well-known solution, the RI logit of [Matějka and McKay \(2015\)](#); see [Maćkowiak, Matějka, and Wiederholt \(2023\)](#) for a review. By Lemma 3 the soft problem can be posed over state-dependent choice probabilities $P(a \mid \theta)$. For finite Ω and $\mathcal{A}$ and $\lambda > 0$, the optimal choice probabilities are

$$P(a \mid \theta) = \frac{P(a) e^{U(a, \theta)/\lambda}}{\sum_b P(b) e^{U(b, \theta)/\lambda}}, \quad (10)$$

where the unconditional choice probabilities $P \in \Delta(\mathcal{A})$ solve

$$V_{\text{soft}}(\lambda) = \max_{P \in \Delta(\mathcal{A})} E\{W_\lambda(\Theta; P)\}, \quad W_\lambda(\theta; P) := \lambda \ln \sum_a P(a) e^{U(a, \theta)/\lambda}. \quad (11)$$

This reduction of the soft problem to a maximisation over P holds on general state and action spaces (Denti, Marinacci, and Montrucchio, 2020, Theorem 1 and Corollary 8); it rests on the variational formula for relative entropy, which their Lemma 3 generalises. Actions with $P(a) = 0$ are never chosen. Choice probabilities (10) are multinomial logit with alternative-specific constants $\lambda \ln P(a)$, and W_λ is the corresponding logsum, or inclusive value, the expected maximum utility of discrete choice theory.⁶ By (6), the expected logsum (11) is the Fenchel conjugate of the VoI frontier. Divided by λ , the maximisation in (11) is a growth-optimal portfolio problem: action a is an asset with gross return $e^{U(a, \theta)/\lambda}$ in state θ , and $P(a)$ is the share of wealth invested in it. Robson, Samuelson, and Steiner (2023, Proposition 2) show that the unconditional choice probabilities of the rational inattention problem are exactly the growth-optimal portfolio, a result mathematically equivalent to Lemma 2 of Matějka and McKay (2015). Write $I(\lambda)$ for the information acquired by a solution of the soft problem at price λ , and P_λ^* for an optimal P in (11).

Proposition 3 (Hard-Constrained Choice). *Let $0 < \bar{I} < \bar{I}^{\text{sat}}$.*

- (i) Demand for information. V_{soft} is convex and non-increasing, $-I(\lambda) \in$

⁶In statistical mechanics (10) is a Gibbs distribution at temperature λ and W_λ is a free energy.

$\partial V_{\text{soft}}(\lambda)$, and $I(\lambda)$ is non-increasing in λ , with $I(\lambda) = -V'_{\text{soft}}(\lambda)$ wherever V_{soft} is differentiable.

- (ii) Shadow price as inverse demand. $\lambda^* \in \partial V(\bar{I})$ if and only if $-\bar{I} \in \partial V_{\text{soft}}(\lambda^*)$.
- (iii) Value. $U(\bar{I}) = \min_{\lambda \geq 0} \{V_{\text{soft}}(\lambda) + \lambda \bar{I}\}$, the minimum is attained exactly at the shadow prices $\lambda^* \in \partial V(\bar{I})$, all of which are positive, and there $U(\bar{I}) = E\{W_{\lambda^*}(\Theta; P_{\lambda^*}^*)\} + \lambda^* \bar{I}$.
- (iv) Behaviour. At every such λ^* the soft problem has a solution that acquires $\bar{I}$ nats, and every such solution solves the hard problem at $\bar{I}$.

For Shannon costs, part (iii) is the counterpart of [Denti, Marinacci, and Montrucchio \(2020, Proposition 11\)](#), which states the duality for the rate distortion function rather than for the value, and part (iv) is the observation of [Maćkowiak, Matějka, and Wiederholt \(2023, Section 2.4.2\)](#) that the solution under a fixed capacity is a logit with an endogenous shadow cost. What the proposition adds is that the minimising prices are exactly the shadow prices $\partial V(\bar{I})$, so that they exist and are positive, and that they are the inverse of the demand for information, including at kinks of V_{soft} . A hard-constrained agent therefore chooses as an RI logit agent whose price of information is not a parameter but the inverse of its own demand for information at $\bar{I}$. Since that price depends on U , μ and $\mathcal{A}$, it differs across decision problems, which is the source of the global non-equivalence noted by [Azrieli \(2021\)](#). A logit model with a single λ fitted across decision problems that share a capacity is therefore misspecified when the constraint is hard, unless the shadow prices of these problems at $\bar{I}$ happen to coincide; conversely, a fixed capacity misrepresents a soft agent whose attention responds to the stakes ([Section 6](#)). In the

Gaussian benchmark this demand curve is given explicitly by Lemma 2(ii). Where demand is strictly decreasing, $\lambda^*(\bar{I})$ is found by bisection on $I(\lambda) - \bar{I}$; at a kink of V_{soft} the budget is attained by mixing two logit solutions.

6. Across Stakes

Theorems 1–3 relate the two formulations problem by problem. When the decision problem changes, they differ in what they hold fixed: the hard formulation holds the capacity $\bar{I}$ fixed, the soft formulation the price λ . Matějka and McKay (2015, p. 282) stress that only the latter lets the agent pay more attention when the stakes are high, and Azrieli (2021, Proposition 6) shows that the solution of the constrained problem does not change when the stakes are scaled. This section shows that, when the stakes of a decision problem are scaled, the information response of every formulation, the two polar ones and every convex cost of capacity between them, follows from the frontier V of the unscaled problem.

A decision problem is faced *at scale* $k > 0$ if its utility is kU ; write V_k and $V_{\text{soft},k}$ for the VoI and the soft-constraint objective of the scaled problem, and $U_k(0) = kU(0)$. Capacity is acquired at a cost $c : [0, \infty) \rightarrow [0, \infty]$ that is convex, non-decreasing and lower semicontinuous with $c(0) = 0$, so that the agent solves the soft problem with cost functional $c(I(p))$ of the remark following Proposition 1. The two formulations are the polar cases: $c(\bar{I}) = \lambda\bar{I}$ is the linear penalty, and $c = 0$ on $[0, \bar{I}_0]$, $c = +\infty$ beyond, is the hard constraint at $\bar{I}_0$. Write

$$\Pi_c(k) := \sup_{\bar{I} \geq 0} \{kV(\bar{I}) - c(\bar{I})\}, \quad \mathcal{I}_c(k) := \arg \max_{\bar{I} \geq 0} \{kV(\bar{I}) - c(\bar{I})\}.$$

We assume that V is bounded, as in the finite case and the Gaussian benchmark, and that c is not identically zero. Since c is convex with $c(0) = 0$, $c(\bar{I})/\bar{I}$ is non-decreasing, so $c(\bar{I}) \rightarrow \infty$; as $kV - c$ is upper semicontinuous, the set $\mathcal{I}_c(k)$ is then non-empty and compact.⁷ Write $\bar{I}_c^+(k)$ and $\bar{I}_c^-(k)$ for its largest and smallest elements, and $\mathcal{I}_\lambda(k)$ and $\bar{I}_\lambda^\pm(k)$ for the linear penalty $c(\bar{I}) = \lambda \bar{I}$ with $\lambda > 0$.

Proposition 4 (Stakes). *Let $k > 0$.*

- (i) *Scaling. $V_k = kV$, $\partial V_k(\bar{I}) = k \partial V(\bar{I})$, and $V_{\text{soft},k}(\lambda) - U_k(0) = k[V_{\text{soft}}(\lambda/k) - U(0)]$.*
- (ii) *Reduction. The soft problem with cost $c(I(p))$ at scale k has value $U_k(0) + \Pi_c(k)$, and a signal structure solves it if and only if its information $I(p)$ lies in $\mathcal{I}_c(k)$ and it solves the hard problem at that budget.*
- (iii) *Response. $\mathcal{I}_c(k)$ is non-decreasing in k in the strong set order, and so are $\bar{I}_c^+(k)$ and $\bar{I}_c^-(k)$. Under the hard constraint the capacity does not respond to k while its shadow price scales by k ; under the linear penalty the information acquired at stake k is the demand at the effective price λ/k , $\mathcal{I}_\lambda(k) = \mathcal{I}_{\lambda/k}(1)$.*
- (iv) *Comparison. Let $\bar{I}_0 \in \mathcal{I}_c(k_0)$ and $\lambda > 0$ with $\lambda \in \partial c(\bar{I}_0) \cap k_0 \partial V(\bar{I}_0)$. Then $\bar{I}_0 \in \mathcal{I}_\lambda(k_0)$, and*

$$\bar{I}_0 \leq \bar{I}_c^+(k) \leq \bar{I}_\lambda^+(k) \quad \text{for } k \geq k_0, \quad \bar{I}_\lambda^-(k) \leq \bar{I}_c^-(k) \leq \bar{I}_0 \quad \text{for } k \leq k_0.$$

⁷Indeed, boundedness of V and $c(\bar{I}) \rightarrow \infty$ imply that $kV(\bar{I}) - c(\bar{I}) \rightarrow -\infty$, so all maximisers lie in some compact interval $[0, R]$. Upper semicontinuity then ensures that the maximum is attained and that the set of maximisers is closed, hence compact.

(v) Identification. *If $\bar{I}^* \in \mathcal{I}_c(k)$ and $\bar{I}^* > 0$, then $\partial c(\bar{I}^*) \cap k \partial V(\bar{I}^*) \neq \emptyset$; in particular $c'(\bar{I}^*) \in k \partial V(\bar{I}^*)$ wherever c is differentiable.*

Part (iii) separates the formulations by the direction of their response to the stakes, and part (iv) orders its size. Raising the stakes never lowers the smallest or the largest optimal capacity. Calibrate a convex cost and a linear penalty so that they select the same capacity $\bar{I}_0$ at a reference stake k_0 , which by part (v) is possible at every $\bar{I}_0 > 0$. As the stakes rise, the convex cost raises capacity by less than the linear penalty, and as they fall it lowers capacity by less; the hard constraint, the limit of an ever more convex cost, does not respond at all. A convex cost of capacity thus lies between the polar cases in a precise sense. In the Gaussian benchmark, a linear penalty gives $\bar{I}^*(k) = \frac{1}{2} \ln(2k\sigma^2/\lambda)$ for $k \geq \lambda/(2\sigma^2)$, so information rises with the logarithm of the stakes; a quadratic cost $c(\bar{I}) = \gamma\bar{I}^2/2$ gives the capacity solving $2k\sigma^2e^{-2\bar{I}} = \gamma\bar{I}$, which rises by $\bar{I}/(2\bar{I} + 1) < \frac{1}{2}$ per unit of $\ln k$, more slowly; a fixed capacity does not respond at all. [Paciello and Wiederholt \(2014, p. 372\)](#) make the same observation in a model of price setting: the more convex the cost of attention, the smaller the increase in attention for a given increase in its benefit.

Under the hard constraint information does not respond to the stakes: by part (iii) the shadow price of capacity rises with k , but the agent has no way of acting on it. This changes once capacity can be traded. Suppose the agent is endowed with a capacity $\bar{I}_0$, such as the time or attention available to him or her, and can buy more at a price $p > 0$, for instance by acquiring data or hiring an analyst. The cost of capacity is then $c(\bar{I}) = p \max\{\bar{I} - \bar{I}_0, 0\}$, which is convex, non-decreasing and continuous with $c(0) = 0$, so that Proposition 4

applies to it. Write V'_+ for the right derivative of V ; by concavity, $V'_+(\bar{I}_0)$ is the smallest element of $\partial V(\bar{I}_0)$.

Corollary 1 (Endowment and a Market for Capacity). *Let $0 < \bar{I}_0 < \bar{I}^{\text{sat}}$, $p > 0$, $c(\bar{I}) = p \max\{\bar{I} - \bar{I}_0, 0\}$ and $\bar{k} = p/V'_+(\bar{I}_0)$.*

- (i) *For $k < \bar{k}$, $\mathcal{I}_c(k) = \{\bar{I}_0\}$: the agent uses the endowment and buys no capacity.*
- (ii) *For $k > \bar{k}$, $\mathcal{I}_c(k) = \mathcal{I}_p(k)$ (the capacities chosen under a linear penalty at the market price p), and all its elements exceed $\bar{I}_0$: the agent buys capacity and acquires the demand at the effective price p/k .*
- (iii) *The threshold stake identifies the shadow price of the endowment: $V'_+(\bar{I}_0) = p/\bar{k}$.*

The corollary nests both formulations. Below the threshold stake the agent behaves as under a hard constraint, as in [Azrieli \(2021, Proposition 6\)](#); above it, as under a linear penalty at price p , as in [Matějka and McKay \(2015\)](#). The stake at which information starts to respond reveals the shadow price of the hard constraint, which in a pure hard formulation has no behavioural counterpart. Compared with a linear Shannon agent facing the same price p , the agent acquires more information at low stakes and the same at high stakes, so that his or her information responds less to the stakes over the whole range. In the Gaussian benchmark, $\bar{k} = p e^{2\bar{I}_0}/(2\sigma^2)$ and $\bar{I}^*(k) = \max\{\bar{I}_0, \frac{1}{2} \ln(2k\sigma^2/p)\}$. The market cost is the convex cost of [Proposition 4](#) with all its curvature at the kink $\bar{I}_0$, where any marginal cost in $[0, p]$ is admissible; this is why the response is first nil and then that of a linear penalty. If the price per bit rises with the quantity bought, the cost is

strictly convex beyond $\bar{I}_0$, the threshold remains, with p the price of the first unit bought, and above it capacity rises by less than under a linear penalty at price p (Proposition 4(iv)).

These patterns are consistent with experimental evidence. [Dean and Ne-
ligh \(2023\)](#) find that accuracy rises with the reward, and conclude that a fixed mutual information constraint is “strongly rejected” in their environment (p. 3423); they also find that subjects are less responsive to incentives than the Shannon model predicts (p. 3418), and their best fitting models allow a more flexible response. A response between the polar cases is what part (iv) predicts for a convex cost of capacity. In other tasks a substantial share of subjects do not respond to the stakes at all (see the discussion in [Azrieli, 2021](#), Section 1), as Corollary 1 predicts for stakes below the threshold.

Remark (Amortisation). Π_c is convex, as a supremum of the affine functions $k \mapsto kV(\bar{I}) - c(\bar{I})$, and for every $\bar{I}^* \in \mathcal{I}_c(k)$ the function indexed by $\bar{I}^*$ supports Π_c at k , so that $V(\bar{I}^*) \in \partial\Pi_c(k)$. Moreover $\Pi_c(k)/k = \sup_{\bar{I}} \{V(\bar{I}) - c(\bar{I})/k\}$ is non-decreasing in k because $c \geq 0$, and strictly increasing at every k at which some $\bar{I}^* \in \mathcal{I}_c(k)$ has $c(\bar{I}^*) > 0$: for $k' > k$, $\Pi_c(k')/k' \geq V(\bar{I}^*) - c(\bar{I}^*)/k' > V(\bar{I}^*) - c(\bar{I}^*)/k = \Pi_c(k)/k$. Information thus behaves as a fixed cost in the scale of the decision problem. The capacity is bought once and used on the whole stake, so the payoff per unit of scale rises towards the informed value as the cost is spread, and the marginal unit of scale earns the full informed value $V(\bar{I}^*)$. This is the source of increasing returns to information identified by [Arrow \(1996, Section IV\)](#): the value of information about rates is proportional to the scale on which it is used while its cost is not,

*so the demand for information and the rate of return both rise with scale.*⁸

Parts (iii) and (iv) turn such observations into a test of the formulations, and part (v) and Corollary 1(iii) into a measurement. The frontier V is determined by the decision problem; given V , the capacity chosen at each stake reveals the marginal cost of capacity at that capacity, so that variation in the stakes traces out c' on the range of capacities it sweeps, without a parametric form. In some settings the capacity is observable directly: information arrives through a channel of known capacity, as in the sender–receiver foundation of the constrained problem by [Le Treust and Tomala \(2019\)](#), or through data whose volume is fixed by a sampling or survey budget. There the frontier alone determines behaviour and its shadow price, and no cost schedule needs to be estimated, whereas a marginal cost of information valid at every information level is difficult to measure.

In the terms of Section 3, the endowment of Corollary 1 is the agent’s processing capacity, priced by its shadow price, and the market supplies capacity at a price per bit, as [Sims \(2006, p. 160\)](#) takes wiring to be priced. The purchased capacity must relax the constraint that binds. Information that passes through both a wiring channel and the agent’s processing is bounded, by the data-processing inequality, by the smaller of the two capacities, so data bought beyond the processing endowment are worthless unless processing capacity is bought too, as when an analyst is hired or the processing is automated.

⁸[Behringer and Belavkin \(2026b\)](#) use this asymmetry for repeated testing against a common hypothesis, a fleet routed on one network and zero-sum games at different stakes, and [Behringer \(2026c\)](#) for monopoly pricing across markets.

Automated decisions make this structure literal. When an algorithm rather than a person maps market data into trades, as increasingly in financial markets ([Hendershott, Jones, and Menkveld, 2011](#)), human processing leaves the chain from the state to the trade, and two constraints remain: the system’s wiring, which [Sims \(2006, p. 160\)](#) already singles out for financial firms active in many markets, and its compute. Both are installed capacities rather than prices per bit. The wiring is a Shannon capacity in the strict sense: a data line delivers at most a fixed number of bits in the time available for a decision, and this bounds the mutual information between the state and the trade. The bound is attained only by an ideal code, so the frontier at the line’s capacity bounds what an actual system achieves from above. Compute is different: like Sims’s costly investigation, the compute an algorithm spends on inference need not be proportional to the bits it delivers. A compute budget is therefore a hard constraint on a cost functional other than mutual information, and [Section 3](#) applies to it when that functional has the properties listed after [Theorem 1](#). In the short run both capacities are fixed across decision problems, as the hard formulation assumes, and observable, as in the settings above; over a longer horizon they can be extended at a price, which is the setting of [Corollary 1](#).

Financial markets thus show the two formulations as complementary regimes of one model. The hard formulation describes installed capacity, human or automated, in the short run and below the threshold stake; the soft formulation, at the price of capacity, describes its extension above that stake; and costly investigation calls for a cost functional other than mutual information, to which the duality applies but bits do not.

7. Conclusion

The conjugate duality between the hard and soft formulations of RI is standard; this paper has traced what it implies. The soft-constraint objective is the Fenchel conjugate of the VoI, the VoI is fully recoverable from it by the biconjugate theorem (Theorem 2), and, given the capacity, the marginal cost of information is the slope of the VoI by the inverse subgradient theorem (Theorem 3). These results hold beyond Shannon costs, for the class of information costs identified in Section 3, which includes the f -divergence analogues of mutual information. A budget-independent price of information requires the VoI to be affine. The rate distortion approach of [Stratonovich \(1975/2020\)](#), combined with the convex duality of [Rockafellar \(1974\)](#), thus provides a natural foundation for RI models with non-linear information costs.

Neither formulation is free of primitives: the hard formulation fixes the capacity and the soft formulation the price, and they part company across decision problems. For changes in the stakes the frontier makes that comparison precise (Proposition 4): scaling the stakes leaves information unchanged under a fixed capacity, lowers the effective price under a linear penalty, and raises capacity under a convex cost of capacity by less than under a linear penalty, and the response identifies the marginal cost of capacity. With an endowment of capacity and a market for more, the shadow price of the hard constraint becomes observable as the stake at which information starts to respond (Corollary 1). The two formulations are thus complementary regimes of one model. Financial markets illustrate both: installed capacity, which automated trading turns into a Shannon capacity in the literal sense, is de-

scribed by the hard formulation in the short run, and its extension at a price by the soft one, while the costs of costly investigation are probably not well measured in bits at all (Sims, 2006, p. 160). Where capacity is installed, observable and cannot be extended over the horizon considered, the hard formulation is the more economical description: the frontier alone determines behaviour and its shadow price, including under changes in the stakes or the prior, and no cost function has to be chosen or estimated.

For Shannon costs the duality has a direct behavioural reading. The conjugate is the expected logsum (inclusive value) of the RI logit, the shadow price of capacity is the inverse of the logit agent’s demand for information, and a hard-constrained agent chooses as a logit agent at that endogenous price (Proposition 3); the Gaussian quadratic benchmark, the workhorse of the applied RI literature (Maćkowiak, Matějka, and Wiederholt, 2023), gives all three in closed form. A logit price estimated from choice data therefore measures the price of information if the constraint is soft and the shadow price of capacity if it is hard, and the response of choices to the stakes tells the two readings apart (Proposition 4).

Three extensions are natural. First, Corollary 1 takes the price of capacity as given; with many agents, such as automated traders buying data lines or compute, the price would be determined in equilibrium together with the information acquired, and the threshold stakes would sort the agents who buy capacity from those who do not. Second, capacity that is fixed in the short run and extended over a longer horizon suggests a dynamic version in which capacity is an investment whose timing is governed by its shadow price. Third, the logit form of Section 5 is specific to Shannon costs; for

other costs in the class of Section 3 the soft problem no longer yields logit choice, and what replaces Proposition 3 there is open.

Appendix A. Proofs

Proof of Theorem 1. Fix $\lambda \geq 0$. No property of the cost functional C beyond $C(p) \in [0, \infty)$ is used. The identity interchanges the order of the suprema over signal structures and over budgets:

$$\begin{aligned} V_{\text{soft},C}(\lambda) &= \sup_p \{E\{U\} - \lambda C(p)\} = \sup_p \sup_{\bar{I} \geq C(p)} \{E\{U\} - \lambda \bar{I}\} \\ &= \sup_{\bar{I} \geq 0} \sup_{p: C(p) \leq \bar{I}} \{E\{U\} - \lambda \bar{I}\} = \sup_{\bar{I} \geq 0} \{U_C(\bar{I}) - \lambda \bar{I}\}. \end{aligned}$$

The first equality is the definition of $V_{\text{soft},C}$; the second uses $\lambda \geq 0$, so that the inner supremum is attained at the smallest admissible budget $\bar{I} = C(p)$; the third takes the supremum over the same set of pairs $(p, \bar{I})$ with $C(p) \leq \bar{I}$ in the opposite order; the fourth uses the definition of U_C , with $\sup \emptyset = -\infty$. $\square$

Proof of Theorem 2. Extend $-U$ to $\mathbb{R}$ by setting $-U(\bar{I}) := +\infty$ for $\bar{I} < 0$.⁹ By (6), $V_{\text{soft}}(\lambda) = \sup_{\bar{I} \geq 0} \{U(\bar{I}) - \lambda \bar{I}\} = (-U)^*(-\lambda)$ for $\lambda \geq 0$, the Fenchel conjugate of $-U$ evaluated at $-\lambda$. For $\lambda < 0$ the right-hand side is $+\infty$, since U is non-decreasing and $-\lambda \bar{I} \rightarrow \infty$ as $\bar{I} \rightarrow \infty$; we extend V_{soft} by $+\infty$ there, so that $(-U)^*(y) = V_{\text{soft}}(-y)$ for all $y \in \mathbb{R}$. By Lemma 4, $-U$ is proper, closed and convex. By Rockafellar (1974, Theorem 5, p. 16), $\text{cl co}(-U) = -U$, so the biconjugate satisfies $(-U)^{**} = -U$. Unfolding the biconjugate in terms of V_{soft} and substituting $\lambda = -y$:

$$(-U)^{**}(\bar{I}) = \sup_{y \in \mathbb{R}} \{y \bar{I} - (-U)^*(y)\} = \sup_{\lambda \in \mathbb{R}} \{-\lambda \bar{I} - V_{\text{soft}}(\lambda)\}.$$

⁹This makes $-U$ a convex function on $\mathbb{R}$, as required by Rockafellar (1974); points $\bar{I} < 0$ then drop out of the conjugate (a negative information budget makes no sense), which reduces to the supremum over $\bar{I} \geq 0$ in Theorem 1. Appendix B shows that this extension arises automatically as the optimal value of the perturbed problem.

Since $V_{\text{soft}}(\lambda) = +\infty$ for $\lambda < 0$, these terms do not contribute. Hence $U(\bar{I}) = \inf_{\lambda \geq 0} \{V_{\text{soft}}(\lambda) + \lambda \bar{I}\}$, and subtracting the constant $U(0)$ gives the statement for V . For $\bar{I} > 0$ the infimum is attained at every $\lambda^* \in \partial V(\bar{I})$, which is non-empty, by Theorem 3 (Proposition 3(iii)). Appendix B derives the same identity from Rockafellar (1974, Theorem 7). $\square$

Proof of Theorem 3. By Lemma 4, $-U$, extended by $+\infty$ for $\bar{I} < 0$ as in the proof of Theorem 2, is a proper closed convex function with non-empty subdifferential at every $\bar{I} > 0$. The subdifferential of the concave function $V = U - U(0)$ defined in Section 2 is $\partial V(\bar{I}) = -\partial(-U)(\bar{I})$, which the constant $U(0)$ does not affect. By Rockafellar (1974, Corollary 12A, p. 35) applied to $\varphi = -U$, $u = \bar{I}$:

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

By (6) and the extension of V_{soft} in the proof of Theorem 2, $(-U)^*(y) = V_{\text{soft}}(-y)$, so $\partial(-U)^*(y) = -\partial V_{\text{soft}}(-y)$ by the chain rule. Translating with $y = -\lambda^*$:

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

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^* , $\partial V_{\text{soft}}(\lambda^*) = \{V'_{\text{soft}}(\lambda^*)\} = \{-\bar{I}\}$, establishing the mutual identification of primal and dual solutions. $\square$

Proof of Proposition 2. By Theorem 3, $\partial V(\bar{I})$ depends only on the frontier V , which is determined by U , μ and $\mathcal{A}$. If a single λ lies in $\partial V(\bar{I})$ for every $\bar{I}$ in an interval $[\bar{I}_1, \bar{I}_2]$, then $V(\bar{I}_2) \leq V(\bar{I}_1) + \lambda(\bar{I}_2 - \bar{I}_1)$ and $V(\bar{I}_1) \leq V(\bar{I}_2) - \lambda(\bar{I}_2 - \bar{I}_1)$, so $V(\bar{I}_2) - V(\bar{I}_1) = \lambda(\bar{I}_2 - \bar{I}_1)$; applied to any two

points of the interval, this makes V affine with slope λ there. Conversely, if V is affine with slope λ on the interval, concavity gives $\lambda \in \partial V(\bar{I})$ at every point of the interval, and by Theorem 3 $-\bar{I} \in \partial V_{\text{soft}}(\lambda)$ for each such $\bar{I}$, so that every budget in the interval attains the supremum in (6) at price λ . $\square$

Proof of Lemma 1. (i) Under quadratic utility the Bayes-optimal action after signal s is $a^*(s) = E\{\Theta \mid S = s\}$, and its conditional expected utility is $-\text{Var}(\Theta \mid S = s)$. Taking expectations gives $E\{U(a^*(S), \Theta)\} = -E\{\text{Var}(\Theta \mid S)\}$. Without a signal the optimal action is μ_0 , so $U(0) = -\sigma^2$.

(ii) Let $p(s \mid \theta)$ satisfy $I(\Theta, S) \leq \bar{I}$ and put $A := a^*(S) = E\{\Theta \mid S\}$. Since A is a function of S , the data-processing inequality (Cover and Thomas, 2006, Theorem 2.8.1) gives $I(\Theta, A) \leq I(\Theta, S) \leq \bar{I}$, so A is feasible in (7) and $E\{U(a^*(S), \Theta)\} = -E\{(A - \Theta)^2\} \leq -D(\bar{I})$. Taking the supremum over signal structures gives $U(\bar{I}) \leq -D(\bar{I})$. Conversely, any channel $q(a \mid \theta)$ with $I(\Theta, A) \leq \bar{I}$ is a feasible signal structure with signal A , and the Bayes-optimal response $E\{\Theta \mid A\}$ satisfies $E\{(E\{\Theta \mid A\} - \Theta)^2\} \leq E\{(A - \Theta)^2\}$. Hence $U(\bar{I}) \geq -E\{(A - \Theta)^2\}$ for every such channel, and taking the infimum gives $U(\bar{I}) \geq -D(\bar{I})$. The expression for V follows from $U(0) = -\sigma^2$. $\square$

Proof of Lemma 2. (i) By Lemma 1(ii), $V(\bar{I}) = \sigma^2 - D(\bar{I})$, and the distortion-rate function of a Gaussian source under squared error is $D(\bar{I}) = \sigma^2 e^{-2\bar{I}}$ (Cover and Thomas, 2006, Theorem 10.3.2), which gives (8). The signal in the lemma attains it: S is Gaussian, $I(\Theta, S) = \frac{1}{2} \ln(1 + \sigma^2/\sigma_\varepsilon^2) = \bar{I}$, and the posterior variance is $\sigma^2 \sigma_\varepsilon^2 / (\sigma^2 + \sigma_\varepsilon^2) = \sigma^2 e^{-2\bar{I}}$. The derivatives follow by differentiation, and $\bar{I}^{\text{sat}} = \infty$ because $V(\bar{I}) < \sigma^2 = E\{\max_a U(a, \Theta)\} - U(0)$ for every finite $\bar{I}$.

(ii) By (6), $V_{\text{soft}}(\lambda) - U(0) = \sup_{\bar{I} \geq 0} \{\sigma^2(1 - e^{-2\bar{I}}) - \lambda\bar{I}\}$, a strictly concave problem. For $\lambda \leq 2\sigma^2$ the first-order condition $2\sigma^2 e^{-2\bar{I}} = \lambda$ gives $\bar{I} = \frac{1}{2} \ln(2\sigma^2/\lambda) \geq 0$; for $\lambda \geq 2\sigma^2$ the derivative is non-positive on $[0, \infty)$ and the maximiser is $\bar{I} = 0$. By the proof of Theorem 1, p solves the soft problem if and only if $(p, I(p))$ attains the supremum over pairs, so this maximiser is $I(\lambda)$, attained by the signal of part (i). Substituting $e^{-2I(\lambda)} = \lambda/(2\sigma^2)$ gives (9).

Remark following the lemma. Let the prior have variance σ^2 and take the same signal $S = \Theta + \varepsilon$. By the maximum entropy property of the Gaussian (Cover and Thomas, 2006, Theorem 8.6.5), applied to $S - E\{S\}$, $I(\Theta, S) = h(S) - h(\varepsilon) \leq \frac{1}{2} \ln(1 + \sigma^2/\sigma_\varepsilon^2) = \bar{I}$, and the linear estimate $\mu_0 + \sigma^2(S - \mu_0)/(\sigma^2 + \sigma_\varepsilon^2)$ has mean squared error $\sigma^2\sigma_\varepsilon^2/(\sigma^2 + \sigma_\varepsilon^2) = \sigma^2 e^{-2\bar{I}}$. Hence $D(\bar{I}) \leq \sigma^2 e^{-2\bar{I}}$, and the bound follows from Lemma 1(ii). $\square$

Proof of Proposition 3. (i) By Lemma 3, $V_{\text{soft}}(\lambda) = \max_{q \in Q} \{L(q) - \lambda I(q)\}$ is a supremum of affine functions of λ with slopes $-I(q) \leq 0$, hence convex and non-increasing. If q_λ solves the soft problem at λ , then for every $\lambda' \geq 0$

$$V_{\text{soft}}(\lambda') \geq L(q_\lambda) - \lambda' I(q_\lambda) = V_{\text{soft}}(\lambda) - (\lambda' - \lambda) I(\lambda),$$

so $-I(\lambda) \in \partial V_{\text{soft}}(\lambda)$. Subgradients of a convex function are non-decreasing, and at a point of differentiability the subgradient is the derivative.

(ii) is Theorem 3. By (i) and (iv), λ^* is therefore a shadow price of the budget $\bar{I}$ if and only if λ^* is a price at which the soft agent demands exactly $\bar{I}$.

(iii) By Theorem 2, $U(\bar{I}) = \inf_{\lambda \geq 0} \{V_{\text{soft}}(\lambda) + \lambda \bar{I}\}$. A point λ minimises this convex function if and only if $-\bar{I} \in \partial V_{\text{soft}}(\lambda)$, that is, by (ii), if and only

if $\lambda \in \partial V(\bar{I})$, a set that is non-empty for $\bar{I} > 0$ and contains only positive prices for $\bar{I} < \bar{I}^{\text{sat}}$ (Section 2). Substituting (11) gives the last formula.

(iv) The set of solutions of the soft problem at λ^* is convex, because $L - \lambda^* I$ is concave on Q , and compact. The subdifferential of a maximum of affine functions is the convex hull of the slopes of the maximisers, so $-\bar{I} \in \partial V_{\text{soft}}(\lambda^*)$ lies between $-I$ evaluated at two solutions; since I is continuous along the segment joining them, some solution acquires $\bar{I}$ nats. If q solves the soft problem at λ^* with $I(q) = \bar{I}$, then $L(q) = V_{\text{soft}}(\lambda^*) + \lambda^* \bar{I} = U(\bar{I})$ by (iii), and q is feasible at $\bar{I}$. $\square$

Proof of Proposition 4. (i) Scaling utility by $k > 0$ leaves the Bayes-optimal rule unchanged, so every signal structure attains k times its expected utility, and $U_k(\bar{I}) = kU(\bar{I})$ for every budget; in particular $U_k(0) = kU(0)$ and $V_k = kV$. The subgradient inequality defining ∂V scales with k . By (6), $V_{\text{soft},k}(\lambda) - U_k(0) = \sup_{\bar{I} \geq 0} \{kV(\bar{I}) - \lambda \bar{I}\} = k \sup_{\bar{I} \geq 0} \{V(\bar{I}) - (\lambda/k) \bar{I}\} = k[V_{\text{soft}}(\lambda/k) - U(0)]$.

(ii) Since c is non-decreasing, $\sup_{\bar{I} \geq I(p)} \{kE\{U\} - c(\bar{I})\} = kE\{U\} - c(I(p))$ for every p . Interchanging the suprema over p and over budgets as in the proof of Theorem 1 gives the value $\sup_{\bar{I}} \{U_k(\bar{I}) - c(\bar{I})\} = U_k(0) + \Pi_c(k)$. A signal structure p attains it if and only if the pair $(p, I(p))$ attains the supremum over pairs, that is, if and only if $I(p) \in \mathcal{I}_c(k)$ and p attains $U_k(I(p))$.

(iii) Points outside the domain of c are never optimal, so the choice may be restricted to that domain, an interval that does not depend on k . There the objective $f(\bar{I}, k) = kV(\bar{I}) - c(\bar{I})$ has increasing differences: for $\bar{I}' > \bar{I}$, $f(\bar{I}', k) - f(\bar{I}, k) = k[V(\bar{I}') - V(\bar{I})] - [c(\bar{I}') - c(\bar{I})]$ is non-decreasing in k because V is non-decreasing. Hence $\mathcal{I}_c(k)$ is non-decreasing in the strong set

order (Milgrom and Shannon, 1994, Theorem 4), and so are its largest and smallest elements. Under the hard constraint at $\bar{I}_0$, the problem at every scale is the hard problem at $\bar{I}_0$, and by (i) its shadow prices are $k \partial V(\bar{I}_0)$. Under the linear penalty, $\arg \max\{kV(\bar{I}) - \lambda \bar{I}\} = \arg \max\{V(\bar{I}) - (\lambda/k)\bar{I}\}$.

(iv) Since $\lambda/k_0 \in \partial V(\bar{I}_0)$, $k_0 V(J) - \lambda J \leq k_0 V(\bar{I}_0) - \lambda \bar{I}_0$ for all $J \geq 0$, so $\bar{I}_0 \in \mathcal{I}_\lambda(k_0)$. Recall that for the convex function c the right derivative c'_+ is non-decreasing, that $\partial c(\bar{I}_0) = [c'_-(\bar{I}_0), c'_+(\bar{I}_0)]$, and that $(c(x) - c(y))/(x - y)$ lies between $c'_+(y)$ and $c'_-(x)$ for $y < x$ in the domain of c (Rockafellar, 1970, Section 24). Let $k \geq k_0$, $x = \bar{I}_c^+(k)$ and $y = \bar{I}_\lambda^+(k)$. By (iii), $x \geq \bar{I}_c^+(k_0) \geq \bar{I}_0$ and $y \geq \bar{I}_\lambda^+(k_0) \geq \bar{I}_0$. Suppose $x > y$. Since y is the largest maximiser of $kV - \lambda \bar{I}$, $kV(x) - \lambda x < kV(y) - \lambda y$, and since x maximises $kV - c$, $kV(x) - c(x) \geq kV(y) - c(y)$; hence $c(x) - c(y) < \lambda(x - y)$. But $c(x) < \infty$ and $\bar{I}_0 \leq y < x$ give $(c(x) - c(y))/(x - y) \geq c'_+(y) \geq c'_+(\bar{I}_0) \geq \lambda$, a contradiction. Let $k \leq k_0$, $x = \bar{I}_c^-(k)$ and $y = \bar{I}_\lambda^-(k)$, so that $x, y \leq \bar{I}_0$ by (iii), and suppose $x < y$. Since y is the smallest maximiser of $kV - \lambda \bar{I}$, $k[V(y) - V(x)] > \lambda(y - x)$, and since x maximises $kV - c$, $c(y) - c(x) \geq k[V(y) - V(x)]$; hence $c(y) - c(x) > \lambda(y - x)$. But $x < y \leq \bar{I}_0$ give $(c(y) - c(x))/(y - x) \leq c'_-(y) \leq c'_-(\bar{I}_0) \leq \lambda$, a contradiction.

(v) Extend V by $-\infty$ to negative budgets. The convex function $c - kV$ attains its minimum at $\bar{I}^*$, so $0 \in \partial(c - kV)(\bar{I}^*)$. The domain of c contains $[0, \bar{I}^*]$ and that of $-V$ is $[0, \infty)$, so their interiors meet, and the subdifferential of the sum is the sum of the subdifferentials (Rockafellar, 1970, Theorem 23.8): $\partial(c - kV)(\bar{I}^*) = \partial c(\bar{I}^*) - k \partial V(\bar{I}^*)$. Hence some element of $\partial c(\bar{I}^*)$ lies in $k \partial V(\bar{I}^*)$; where c is differentiable, $\partial c(\bar{I}^*) = \{c'(\bar{I}^*)\}$. $\square$

Proof of Corollary 1. On $[0, \bar{I}_0]$, $kV - c = kV$, which is strictly increasing

because $\bar{I}_0 < \bar{I}^{\text{sat}}$ (Section 3). On $[\bar{I}_0, \infty)$, $kV(\bar{I}) - c(\bar{I}) = kV(\bar{I}) - p\bar{I} + p\bar{I}_0$, and concavity gives $kV(\bar{I}) - c(\bar{I}) \leq kV(\bar{I}_0) + [kV'_+(\bar{I}_0) - p](\bar{I} - \bar{I}_0)$.

(i) If $k < \bar{k}$, then $kV'_+(\bar{I}_0) < p$, so every $\bar{I} \neq \bar{I}_0$ yields strictly less than $\bar{I}_0$.

(ii) If $k > \bar{k}$, the concave functions $kV - c$ and $kV - p\bar{I}$ both have positive right derivative $kV'_+(\bar{I}_0) - p$ at $\bar{I}_0$, so neither attains its maximum on $[0, \bar{I}_0]$. On $(\bar{I}_0, \infty)$ they differ by the constant $p\bar{I}_0$, so they have the same maximisers, and $\mathcal{I}_c(k) = \mathcal{I}_p(k)$.

(iii) is the definition of $\bar{k}$, and $V'_+(\bar{I}_0) > 0$ because V is concave and strictly increasing on $[0, \bar{I}^{\text{sat}}]$. In the Gaussian benchmark, $V'(\bar{I}) = 2\sigma^2 e^{-2\bar{I}}$ by Lemma 2, which gives $\bar{k}$, and $\mathcal{I}_p(k) = \{\frac{1}{2} \ln(2k\sigma^2/p)\}$ for $k > \bar{k}$. $\square$

Appendix B. Conjugate Duality

This appendix derives Theorems 1 and 2 from the conjugate duality of Rockafellar (1974, Section 4), and in doing so establishes the concavity and closedness of the frontier on which Theorems 2 and 3 rely. Since the framework is less familiar to economists, we proceed in small steps.

Appendix B.1. Rockafellar's Framework

Rockafellar's duality theory is formulated for minimisation problems. A problem of minimising over a variable x is embedded in a family of perturbed problems indexed by a parameter u , through a function $F(x, u)$ with values in $\mathbb{R} \cup \{+\infty\}$, where the value $+\infty$ encodes infeasibility. Following the later literature, we call F the *perturbation function*. Three functions are attached to F : the optimal value as a function of the perturbation,

$$\varphi(u) := \inf_x F(x, u);$$

the Lagrangian, in which the multiplier y prices the perturbation,

$$K(x, y) := \inf_u \{F(x, u) + \langle u, y \rangle\};$$

and the dual function

$$g(y) := \inf_x K(x, y).$$

Here u ranges over a linear space of perturbations and y over a linear space of multipliers, and $\langle u, y \rangle$ is a bilinear pairing between the two, which assigns to a perturbation u its value at the multiplier y . In the finite-dimensional case relevant here, $u, y \in \mathbb{R}^m$ and $\langle u, y \rangle = \sum_{i=1}^m u_i y_i$ is the ordinary inner product; below, the perturbation is a single information budget, $m = 1$, and $\langle u, y \rangle = uy$ is simply the budget multiplied by its price. The term $\langle u, y \rangle$ is therefore the cost, at prices y , of the perturbation u .

The infimum over u has a direct interpretation. For a given choice x , it selects the perturbation that is cheapest at prices y among those under which x is feasible, since infeasible u carry $F(x, u) = +\infty$. In the familiar case of an inequality constraint, $F(x, u) = f(x)$ if $h(x) \leq u$ and $+\infty$ otherwise, the infimum is attained at the binding perturbation $u = h(x)$ when $y \geq 0$, giving $K(x, y) = f(x) + y h(x)$, the textbook Lagrangian with multiplier y on the constraint $h(x) \leq 0$. When $y < 0$ the infimum is $-\infty$: a negative price rewards the agent for every unit of slack, and slack can be increased without bound. The multiplier thus acts as a shadow price, and the sign restriction $y \geq 0$ familiar from the Kuhn–Tucker conditions emerges from the construction rather than being imposed.

By the definitions, $g(y) = \inf_{x,u} \{F(x, u) + \langle u, y \rangle\}$: the dual function is the optimal value when the perturbation u is no longer fixed but chosen together

with x , at the cost given by the pairing $\langle u, y \rangle$. It provides lower bounds on the optimal value of every perturbed problem, since $g(y) \leq F(x, u) + \langle u, y \rangle$ for all x and u , hence

$$g(y) \leq \varphi(u) + \langle u, y \rangle \quad \text{for all } u, y,$$

which is weak duality. The results below state when these bounds are tight, so that φ can be recovered from g .

Theorem 7 of [Rockafellar \(1974\)](#) relates g to φ . Written out without the notation of concave conjugates, it states that for all y and u

$$g(y) = \inf_u \{ \varphi(u) + \langle u, y \rangle \}, \tag{B.1}$$

$$\sup_y \{ g(y) - \langle u, y \rangle \} = (\text{cl co } \varphi)(u), \tag{B.2}$$

where $\text{cl co } \varphi$ is the closed convex hull of φ , the largest closed convex function lying below φ . In Rockafellar's notation, (B.1) is $g = (-\varphi)^*$ and (B.2) is $-g^* = \text{cl co } \varphi$, with conjugates taken in the concave sense. Identity (B.1) requires no assumption: substituting the definitions, $g(y) = \inf_x \inf_u \{ F(x, u) + \langle u, y \rangle \}$, and taking the infimum over x first gives (B.1). Identity (B.2) is the biconjugate theorem. Two further facts are used below: by [Rockafellar \(1974, Theorem 1\)](#), φ is convex whenever F is jointly convex in (x, u) ; and a proper, closed, convex function coincides with its closed convex hull, in which case (B.2) recovers φ from g .

Appendix B.2. Reduction to Action Recommendations

Signal structures $p(s \mid \theta)$ are not a convenient choice variable for this framework: the signal space is unrestricted, and expected utility depends on p non-linearly, through the Bayes-optimal rule a^* . We therefore let the

agent choose directly which action to take in each state, that is, a signal structure whose signals are action recommendations. Let Q denote the set of such channels: conditional probabilities $q(a \mid \theta)$ of recommending action a in state θ , which form a compact convex subset of $\mathbb{R}^{A \times \Omega}$ (a finite product of probability simplices). They are the state-dependent choice probabilities $P(a \mid \theta)$ of Section 5. For $q \in Q$ let

$$L(q) := \sum_{\theta} \mu(\theta) \sum_a q(a \mid \theta) U(a, \theta)$$

be the expected utility of following a recommended action drawn from q , and let $I(q)$ be the mutual information between Θ and the recommendation, i.e. $I(p)$ in (1) evaluated at the signal structure $p = q$. The objective L is linear in q , which is what the convexity argument below requires.

Lemma 3 (Recommendations). *For every $\bar{I} \geq 0$ and $\lambda \geq 0$,*

$$U(\bar{I}) = \max_{q \in Q} \{L(q) : I(q) \leq \bar{I}\}, \quad V_{\text{soft}}(\lambda) = \max_{q \in Q} \{L(q) - \lambda I(q)\}.$$

Proof. A signal structure $p(s \mid \theta)$ together with the Bayes-optimal rule $a^*(s)$ induces the channel $q(a \mid \theta) = P\{a^*(S) = a \mid \theta\}$, with $L(q) = E\{U(a^*(S), \Theta)\}$ and, since $a^*(S)$ is a function of S , $I(q) \leq I(p)$ by the data-processing inequality (Cover and Thomas, 2006, Theorem 2.8.1). Hence q is feasible whenever p is, and attains at least the same value in both problems; for the soft problem this uses $\lambda \geq 0$. Conversely, each $q \in Q$ is itself a signal structure, with the recommendation as the signal, and the Bayes-optimal response to it does at least as well as following it, at the same mutual information. The maxima are attained because Q is compact, L is linear, I is continuous on Q because the state and action spaces are finite, and the feasi-

ble sets are non-empty: an uninformative channel, with $q(a \mid \theta)$ independent of θ , has $I(q) = 0$. $\square$

Appendix B.3. From the Perturbation Function to Theorems 1 and 2

Step 1 (Perturbation function). Take as variable $x = q \in \mathbb{R}^{A \times \Omega}$ and as perturbation the information budget $u = \bar{I} \in \mathbb{R}$, and set

$$F(q, \bar{I}) := -L(q) + \delta[q \in Q, I(q) \leq \bar{I}], \quad (\text{B.3})$$

where $\delta[\cdot]$ is 0 when the condition holds and $+\infty$ otherwise. The minus sign turns the maximisation of expected utility into a minimisation, and the indicator imposes the capacity constraint as a hard constraint. This is the perturbation function of the remark following Theorem 1, with the signal structure replaced by the recommendation channel, which by Lemma 3 loses nothing.

Step 2 (Optimal value). For $\bar{I} \geq 0$, Lemma 3 gives

$$\varphi(\bar{I}) = \inf_q F(q, \bar{I}) = -\max_{q \in Q} \{L(q) : I(q) \leq \bar{I}\} = -U(\bar{I}).$$

For $\bar{I} < 0$ no channel satisfies $I(q) \leq \bar{I}$, since mutual information is non-negative, so $\varphi(\bar{I}) = +\infty$. Thus φ is exactly the extension of $-U$ by $+\infty$ used in the proofs of Theorems 2 and 3; the perturbation framework produces it automatically.

Step 3 (Lagrangian). Fix $q \in Q$. Since $F(q, \bar{I}) = -L(q)$ for $\bar{I} \geq I(q)$ and $+\infty$ otherwise,

$$K(q, \lambda) = \inf_{\bar{I} \geq I(q)} \{-L(q) + \lambda \bar{I}\} = \begin{cases} -[L(q) - \lambda I(q)], & \lambda \geq 0, \\ -\infty, & \lambda < 0. \end{cases}$$

For $\lambda \geq 0$ the infimum is attained at the smallest admissible budget, $\bar{I} = I(q)$; for $\lambda < 0$ it is approached as $\bar{I} \rightarrow \infty$. For $q \notin Q$, $F(q, \cdot) = +\infty$ and so $K(q, \lambda) = +\infty$. For $\lambda \geq 0$, K is minus the soft-constraint objective restricted to recommendation channels, so Rockafellar's multiplier y is the information price λ .

Step 4 (Dual function). Minimising over q and using Lemma 3,

$$g(\lambda) = \inf_q K(q, \lambda) = \begin{cases} -V_{\text{soft}}(\lambda), & \lambda \geq 0, \\ -\infty, & \lambda < 0. \end{cases}$$

Steps 3 and 4 together establish $g = -V_{\text{soft}}$ on $[0, \infty)$, with $g = -\infty$ for $\lambda < 0$.

Step 5 (Theorem 1). Inserting Step 2 into (B.1), and noting that the points $\bar{I} < 0$ contribute $+\infty$, for $\lambda \geq 0$

$$g(\lambda) = \inf_{\bar{I} \geq 0} \{-U(\bar{I}) + \lambda \bar{I}\} = -\sup_{\bar{I} \geq 0} \{U(\bar{I}) - \lambda \bar{I}\}.$$

Comparing with Step 4 gives $V_{\text{soft}}(\lambda) = \sup_{\bar{I} \geq 0} \{U(\bar{I}) - \lambda \bar{I}\}$, which is (6), the Shannon case of Theorem 1. Identity (B.1) is an interchange of infima, which is also the content of the direct proof of Theorem 1 in Appendix A. For this step the reduction of Lemma 3 is not needed: taking the signal structure $p(s | \theta)$ itself as the variable, with $C(p)$ in place of $I(q)$, Steps 1–5 go through unchanged for any cost functional, which is why Theorem 1 holds in that generality.

Step 6 (Convexity and closedness). This is the step that uses properties of mutual information.

Lemma 4 (Concavity and Continuity of the VoI). *The VoI $V(\bar{I}) = U(\bar{I}) - U(0)$ is non-decreasing, concave and continuous on $[0, \infty)$. Equivalently, the function φ of Step 2 is proper, closed and convex.*

Proof. By Step 2, $\varphi(\bar{I}) = -U(\bar{I}) = -V(\bar{I}) - U(0)$ for $\bar{I} \geq 0$, and $\varphi(\bar{I}) = +\infty$ for $\bar{I} < 0$. Each property of φ is therefore a property of the VoI with the sign reversed: φ convex means V concave, φ proper means V finite on $[0, \infty)$, and φ closed (lower semicontinuous) means that the VoI cannot jump upwards at any budget, that is, $V(\bar{I})$ is at least the limit of V along any sequence of budgets converging to $\bar{I}$.

Concavity. The term $-L(q)$ is linear, hence convex. The indicator in (B.3) is convex in $(q, \bar{I})$ because the set $G = \{(q, \bar{I}) : q \in Q, I(q) \leq \bar{I}\}$ is convex: Q is convex, and mutual information is convex in the channel for a fixed prior (Cover and Thomas, 2006, Theorem 2.7.4), so for $(q_1, \bar{I}_1), (q_2, \bar{I}_2) \in G$ and $t \in [0, 1]$

$$I(tq_1 + (1-t)q_2) \leq tI(q_1) + (1-t)I(q_2) \leq t\bar{I}_1 + (1-t)\bar{I}_2.$$

Hence F is jointly convex, φ is convex by Rockafellar (1974, Theorem 1), and the VoI is concave. In economic terms: mixing the optimal channels for two budgets is feasible for the averaged budget and yields the averaged expected utility.

Finiteness. $\varphi > -\infty$ everywhere because L is bounded on the compact set Q , and $\varphi(\bar{I}) < \infty$ for $\bar{I} \geq 0$ because uninformative channels are feasible. Hence the VoI is finite on $[0, \infty)$.

No upward jumps (closedness). Let $\bar{I}_n \rightarrow \bar{I}$ with $\liminf_n \varphi(\bar{I}_n) = \ell < \infty$, and pass to a subsequence along which $\varphi(\bar{I}_n) \rightarrow \ell$; along it $\bar{I}_n \geq 0$, since

$\varphi = +\infty$ on $(-\infty, 0)$. By Lemma 3 there are $q_n \in Q$ with $I(q_n) \leq \bar{I}_n$ and $-L(q_n) = \varphi(\bar{I}_n)$. By compactness of Q a further subsequence converges to some $\bar{q} \in Q$, and by continuity of I and L , $I(\bar{q}) = \lim_n I(q_n) \leq \bar{I}$ and $-L(\bar{q}) = \ell$. Hence $\varphi(\bar{I}) \leq \ell$, so φ is lower semicontinuous and the VoI cannot jump upwards.

Continuity of the VoI. A concave function that is finite on an interval is continuous at every interior point of that interval, here on $(0, \infty)$, so a break could occur only at the end point $\bar{I} = 0$. There a break would mean that the VoI jumps upwards immediately after zero, as if an arbitrarily small capacity were worth a discrete amount. The previous step rules this out: $\limsup_{\bar{I} \downarrow 0} V(\bar{I}) \leq V(0) = 0$, while $V(\bar{I}) \geq 0$ because the VoI is non-decreasing, the feasible set growing with $\bar{I}$. Hence the VoI is continuous on $[0, \infty)$. For a concave VoI that is finite on $[0, \infty)$, closedness and continuity are thus the same requirement, which is why the main text states the condition for Theorems 2 and 3 as continuity. $\square$

The proof uses the three properties of mutual information listed in the remark following Theorem 1: the data-processing inequality (in Lemma 3), convexity in the channel, and continuity, of which only lower semicontinuity is needed.

Step 7 (Theorem 2). By Lemma 4, $\text{cl co } \varphi = \varphi$, since φ is already convex and lower semicontinuous (see footnote 4). Inserting Step 4 into (B.2), and noting that the terms with $\lambda < 0$ equal $-\infty$, for $\bar{I} \geq 0$

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

Multiplying by -1 , which turns the supremum into an infimum,

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

which is Theorem 2. The perturbation route also shows why only non-negative prices appear: a negative price would pay the agent for every unit of capacity held, and since capacity can be held without being used, the dual function is $-\infty$ there (Step 3). Theorem 3 then follows from Lemma 4 and Rockafellar (1974, Corollary 12A), as in Appendix A.

References

- Arrow, Kenneth J. 1996. “The Economics of Information: An Exposition.” *Empirica* 23 (2): 119–28.
- Azrieli, Yaron. 2021. “Constrained versus Unconstrained Rational Inattention.” *Games* 12 (1): 3.
- Behringer, Stefan. 2026a. “Value of Information Based Risk Measures: Theory and Applications.” mimeo.
- Behringer, Stefan. 2026b. “Pricing Information: Value of Information, Convex Duality, and the Cost Function in Rational Inattention.” mimeo.
- Behringer, Stefan. 2026c. “How Much Should a Monopolist Know? Information Capacity, Scale, and the Informational Structure of the Firm.” mimeo.
- Behringer, Stefan, and Roman V. Belavkin. 2023. “The Value of Information and Circular Settings.” arXiv:2303.16126 [econ.TH].
- Behringer, Stefan, and Roman V. Belavkin. 2025. “The Value of Information in Economic Contexts.” *Physical Sciences Forum* 12 (1): 6.

- Behringer, Stefan, and Roman V. Belavkin. 2026a. “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.
- Behringer, Stefan, and Roman V. Belavkin. 2026b. *The Value of Information in Economics and Decision Theory: A Continuous Framework from Shannon to Stratonovich*. Book, forthcoming.
- Belavkin, Roman V., Panos Pardalos, and Jose Principe. 2024. “Value of Information in the Mean-Square Case and its Application to the Analysis of Financial Time-Series Forecast.” arXiv:2410.01831 [q-fin.ST].
- Bloedel, Alexander W., Tommaso Denti, and Luca Pomatto. 2025. “Modeling Information Acquisition via f -Divergence and Duality.” arXiv:2510.03482 [econ.TH].
- Caplin, Andrew, Mark Dean, and John Leahy. 2022. “Rationally Inattentive Behavior: Characterizing and Generalizing Shannon Entropy.” *Journal of Political Economy* 130 (6): 1676–1715.
- Cover, Thomas M., and Joy A. Thomas. 2006. *Elements of Information Theory*. 2nd ed. Hoboken: Wiley.
- Dean, Mark, and Nathaniel Neligh. 2023. “Experimental Tests of Rational Inattention.” *Journal of Political Economy* 131 (12): 3415–61.
- 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.

- Denti, Tommaso, Massimo Marinacci, and Aldo Rustichini. 2022. “Experimental Cost of Information.” *American Economic Review* 112 (9): 3106–23.
- 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.
- Hansen, Lars Peter, and Thomas J. Sargent. 2001. “Robust Control and Model Uncertainty.” *American Economic Review* 91 (2): 60–66.
- Hansen, Lars Peter, and Thomas J. Sargent. 2008. *Robustness*. Princeton: Princeton University Press.
- Hébert, Benjamin, and Michael Woodford. 2021. “Neighborhood-Based Information Costs.” *American Economic Review* 111 (10): 3225–55.
- Hendershott, Terrence, Charles M. Jones, and Albert J. Menkveld. 2011. “Does Algorithmic Trading Improve Liquidity?” *Journal of Finance* 66 (1): 1–33.
- De Lara, Michel, and Olivier Gossner. 2020. “Payoffs-Beliefs Duality and the Value of Information.” *SIAM Journal on Optimization* 30 (1): 464–489.
- Le Treust, Maël, and Tristan Tomala. 2019. “Persuasion with Limited Communication Capacity.” *Journal of Economic Theory* 184: 104940.
- Maćkowiak, Bartosz, Filip Matějka, and Mirko Wiederholt. 2023. “Rational Inattention: A Review.” *Journal of Economic Literature* 61 (1): 226–73.

- 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.
- Milgrom, Paul, and Chris Shannon. 1994. “Monotone Comparative Statics.” *Econometrica* 62 (1): 157–80.
- De Oliveira, Henrique, Tommaso Denti, Maximilian Mihm, and Kemal Ozbek. 2017. “Rationally Inattentive Preferences and Hidden Information Costs.” *Theoretical Economics* 12: 621–54.
- Paciello, Luigi, and Mirko Wiederholt. 2014. “Exogenous Information, Endogenous Information, and Optimal Monetary Policy.” *Review of Economic Studies* 81 (1): 356–88.
- Pomatto, Luciano, Philipp Strack, and Omer Tamuz. 2023. “The Cost of Information: The Case of Constant Marginal Costs.” *American Economic Review* 113 (5): 1360–93.
- Robson, Arthur J., Larry Samuelson, and Jakub Steiner. 2023. “Decision Theory and Stochastic Growth.” *American Economic Review: Insights* 5 (3): 357–76.
- Rockafellar, R. Tyrrell. 1970. *Convex Analysis*. Princeton: Princeton University Press.
- 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.

- Shannon, Claude E. 1948. “A Mathematical Theory of Communication.” *Bell System Technical Journal* 27 (3): 379–423.
- Sims, Christopher A. 2003. “Implications of Rational Inattention.” *Journal of Monetary Economics* 50 (3): 665–90.
- Sims, Christopher A. 2006. “Rational Inattention: Beyond the Linear-Quadratic Case.” *American Economic Review* 96 (2): 158–63.
- 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.
- Van Nieuwerburgh, Stijn, and Laura Veldkamp. 2010. “Information Acquisition and Under-Diversification.” *Review of Economic Studies* 77 (2): 779–805.