Rational Inattention and the Radner–Stiglitz Nonconcavity

Stefan Behringer

Value of Information Group, Universität Bielefeld, Germany

Abstract

Radner and Stiglitz (1984) show that a little information is never worth its cost, and note that the conclusion turns on the index in which informativeness is measured, reindexing an example of their own so that it appears to fail. Arrow (1985) sets two decision problems against the theorem and diagnoses the hypothesis responsible: indexed by the Shannon rate of transmission, or by posterior precision, the likelihood is not differentiable at the null. We generalise the mechanism behind that diagnosis and convert it into a criterion in terms of capacity: the mutual information between state and signal. Mutual information is second order in any index that resolves the likelihood differentiably, with coefficient equal to the average conditional variance of the first-order perturbation, so a cost linear in such an index has infinite marginal cost per nat. In capacity units the slope of the gross value of information at the origin is the largest ratio of payoff gain to Kullback–Leibler divergence across posteriors, the value itself being concave. A little information is worth its cost exactly when $C'(0^+) < V'(0^+)$. Arrow's logarithmic portfolio problem is an instance of that inequality; the all-or-nothing demand he found unreasonable in it is the case of equality; and the cost specifications the Rational Inattention literature has since adopted satisfy it, outright under the capacity constraint of Sims (2003) and below a threshold price of information otherwise. The increasing returns to information and the discontinuous demand that Radner and Stiglitz derive disappear for these specifications.

Keywords: value of information, Rational Inattention, mutual information, nonconcavity

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

JEL: D80, D81, D83

Highlights

- The index dependence noted by Radner–Stiglitz and Arrow becomes a slope comparison.
- Capacity is mutual information in nats; the cost per nat is unbounded in a smooth index.
- The slope of the value at zero capacity is computable from the primitives.
- A little information pays iff marginal cost at zero is below marginal value.
- Arrow's portfolio optimum and Rational Inattention costs are instances.

1. Introduction

Radner and Stiglitz (1984) prove that the marginal value of information is non-positive at an information structure that conveys nothing, and strictly negative whenever the marginal cost of information is strictly positive. Their value is net of cost, and the force of the result lies in its gross component: in the index by which they parametrise informativeness, the gross value has zero derivative at the null, so that a marginal cost, however small, goes unoffset. A little information is never worth buying. They observe that the conclusion depends on that index, reindexing their own linear prediction example so that it appears to fail and resolving the appearance by noting that their differentiability hypothesis then fails too. Two implications follow, and we take up both below. If any amount of information has positive net value then, the first increment having negative net value, the value function must rise after falling and so cannot be concave: there are increasing returns over some range. The optimum then jumps across that range rather than moving through it, so demand is not continuous in cost. Chade and Schlee (2002) extend the theorem to general state and signal spaces, separate the hypotheses bearing on the information structure from those bearing on the decision maker, and ask directly whether the nonconcavity is an artifact of the units in which information is counted. Their Corollary 4 answers that it is

not: if $V - C$ fails to be quasiconcave, it fails under every strictly increasing reparametrisation, the reindexing carrying the cost with it. Proposition 2 does not reindex: it changes the cost, charging for nats of mutual information rather than for their index. What is at issue here is accordingly not which index to use but which cost systems are convex in that measure, a question their ordinal result leaves open. Neither they nor De Lara and Gilotte (2007), whose tight sufficient condition bears on the information structure alone, had occasion to consider the cost systems that the Rational Inattention literature was about to adopt.

Arrow (1985) sets two decision problems against the theorem. The first is the portfolio model of Arrow (1971), which is also the model of Section 4 of Radner and Stiglitz (1984) themselves: an investor bets his entire resources on a finite set of states through elementary securities, and may install a channel emitting messages S at a cost of c per unit of its rate of transmission $R = H(X) - E_S\{H(X \mid S)\}$. The rate of transmission is the expected reduction in Shannon entropy and hence the mutual information $I(X, S)$ between state and signal, which is the quantity we call capacity below and measure in nats. In the notation of Arrow (1985), Arrow (1971) shows that the gain in maximum expected utility due to the messages equals R when the utility of wealth is logarithmic. With initial resources A the expected payoff under the optimal portfolio is then $E\{\ln r_x\} + \ln(A - cR) - H(X) + R$, where r_x is the gross return on the security paying in state x, which is concave in R and maximised at $R^* = (A - c)/c > 0$ whenever $c < A$.

In the second problem a normal state is observed through a sample whose cost is proportional to its size and therefore linear in posterior precision h; the agent minimises $h^{-1} + ch$ and chooses $h^* = c^{-1/2}$. In both the value of information is concave in the amount and the optimum is interior. Arrow notes that the Radner–Stiglitz theorem is correct, so one of its hypotheses must fail in his examples, and identifies it (Arrow, 1985, p. 306): what is parametrised is the family of experiments itself, indexed by the rate of transmission R in the first problem and by the posterior precision h in the second, and along either index the conditional density $p_t(s \mid x)$ has infinite derivative in the index at $t = 0$.

The Rational Inattention literature that followed Sims (2003) takes for granted the behaviour Arrow found in his two examples. Agents there acquire strictly positive and typically small quantities of information, charged by mutual information or by an increasing transformation of it (Matějka and McKay, 2015; Denti, 2022; Maćkowiak et al., 2023), and the value functions

obtained are concave with interior optima. The two literatures seldom cite one another, and where they do the disagreement is set down to a difference of assumptions and left there. Arrow's remark is the link. It was made for two examples in a discussion of the firm, and it names a hypothesis without saying which decision problems, and which cost schedules for information, fall on either side of it.

We prove four things. Lemma 1: measured in capacity, the gross value of information is concave, with slope at the origin equal to the largest ratio of payoff gain to divergence across posteriors. The concavity is known; it is the slope that the criterion below needs, and the formula makes it computable from preferences and the prior alone. Lemma 2: in any family of experiments indexed so that the likelihood moves differentiably, mutual information moves at the square of that rate. The coefficient is the average conditional variance of the first-order perturbation, and it is what fixes the rate at which the cost per nat diverges at the origin. The infinite derivative Arrow reports in the conditional density is then the chain rule. Proposition 1: for any cost smooth in that index with strictly positive marginal cost at the null, which is the charge Radner and Stiglitz (1984, p. 38) impose, the marginal cost per nat is unbounded; what the index registers as a flat gross value against a finite marginal cost, capacity registers as a positive marginal value against an infinite one. Then Proposition 2 shows that a little information pays exactly when $C'(0^+) < V'(0^+)$. Arrow's own $c < A$ is that inequality in his portfolio problem, and the cost specifications of Matějka and McKay (2015) and Pomatto et al. (2023) have finite marginal cost at zero and so satisfy it below a threshold price per nat, while the capacity constraint of Sims (2003) has $C'(0^+) = 0$ and satisfies it outright. Measuring information by mutual information therefore does not make the Radner–Stiglitz conclusion disappear, that conclusion being about net value; it turns it from a theorem into the comparison of slopes that Proposition 2 makes. Section 5 draws the consequences.

2. The value of information in capacity units

A decision maker holds prior μ over a state $\theta \in \Theta$, chooses an action $a \in A$, and receives a bounded measurable payoff $u(a, \theta)$. An experiment sends a signal whose distribution depends on θ and so induces a distribution τ over posteriors π which averages back to the prior, $E_\tau\{\pi\} = \mu$.[1] Write S for the signal and D_{KL} for Kullback–Leibler divergence. The *capacity* of the experiment is

$$\kappa \;:=\; E_\tau\{D_{\mathrm{KL}}(\pi\|\mu)\} \;=\; I(\Theta, S) \tag{1}$$

nats, where $I(\Theta, S)$ denotes the mutual information between state and signal; the second equality is an identity.[2] The state is written X in the two examples of Section 1, following Arrow. Write $\varphi(\pi) := \sup_a \int u(a, \theta)\mathrm{d}\pi(\theta)$ for the highest expected payoff attainable at π, which is convex and bounded, and

$$V(\kappa) \;:=\; \sup\{E_\tau\{\varphi(\pi)\} - \varphi(\mu) \;:\; E_\tau\{D_{\mathrm{KL}}(\pi\|\mu)\} \le \kappa\} \tag{2}$$

for the *gross* value of information at capacity κ, the constraint being (1). Let $C(\kappa)$ be a cost with $C(0) = 0$, increasing and convex, so net value is $V(\kappa) - C(\kappa)$. Keeping the two apart matters, because the theorem of Radner and Stiglitz (1984) does not: cost enters there through the payoff itself, which is assumed to fall as the information structure becomes more informative, so no separate schedule is subtracted. Their value function is therefore already net, and their theorem is a statement about $V - C$ where Lemma 1 below is a statement about V.[3]

The separation has a consequence worth recording at the outset. The infinite derivative Arrow finds in the conditional density can surface in either component once the two are held apart and both are expressed in capacity. Under Assumption 1 below it surfaces entirely in the cost, a charge linear in

[1] That property is called Bayes plausibility by Kamenica and Gentzkow (2011), and conversely every τ with it is induced by some experiment, which is what permits (2) to be posed as a problem in τ and allows the distributions constructed in the proof of Lemma 1 to be read back as experiments the decision maker can buy.

[2] Write $I(\Theta, S) = E_{\theta,s}\log\{p(\theta, s)/p(\theta)p(s)\}$. Take expectation over θ. The inner term is $E_{\theta|s}\log\{p(\theta \mid s)/p(\theta)\} = D_{\mathrm{KL}}(\pi_s\|\mu)$, so $I(\Theta, S) = E_\tau\{D_{\mathrm{KL}}(\pi\|\mu)\}$ by Bayes' rule alone.

[3] The separation itself is Chade and Schlee (2002, eq. (2))'s, whose gross value excludes the cost of acquisition and who formulate the choice of information as the maximisation of $V - C$ over the index; what differs here is the coordinate in which C is written, and that V is an envelope over experiments rather than a value along a family.

a smoothly resolving index having infinite marginal cost per nat; when that assumption fails it surfaces in the benefit instead and for a different reason, the gross value acquiring infinite slope at the origin through a kink in f at the prior rather than through any change of variable.

Fix a_μ optimal at the prior and set $\ell(\pi) := \int u(a_\mu, \theta)\mathrm{d}\pi(\theta)$, an affine function with $\ell \leq \varphi$ and $\ell(\mu) = \varphi(\mu)$. Write $f := \varphi - \ell \geq 0$, so that $f(\mu) = 0$ and, by Bayes plausibility and affineness of ℓ, the objective in (2) is $E_\tau\{f(\pi)\}$.

Lemma 1. *The function V defined in (2) is nondecreasing and concave on $[0, \infty)$ with $V(0) = 0$, so that $V'(0^+)$ exists in $[0, +\infty]$ and*

$$V'(0^+) \;=\; \sup_{\kappa > 0} \frac{V(\kappa)}{\kappa} \;=\; \sup_{\pi \neq \mu} \frac{f(\pi)}{D_{\mathrm{KL}}(\pi \| \mu)}. \tag{3}$$

The proof is in Appendix A.

The concavity is not new. Whitmeyer (2025, Prop. 2.3, Cor. 2.4) obtains it, together with strict positivity of the marginal value at zero, for an amount of information of the form $\phi\left[E_\tau\{c(\pi, \mu)\}\right]$ with c a strictly convex divergence and ϕ convex, increasing and zero at the origin, acquired efficiently; (1) is that measure with $c = D_{\mathrm{KL}}$ and ϕ the identity, where his conclusion is weak concavity. What Lemma 1 adds is (3), and Proposition 2 needs it: the criterion there compares $V'(0^+)$ with a marginal cost, so a positive sign is not enough and the slope must be available from the primitives, which (3) makes it, from f and μ alone.

In capacity units the positivity is immediate: by the first equality in (3) the slope is $\sup_\kappa V(\kappa)/\kappa$, so it is strictly positive whenever information is worth anything at any capacity at all, whatever the state space and the prior. One experiment makes the mechanism visible. With finitely many states, revealing θ with probability ε and saying nothing otherwise has capacity $\varepsilon H(\Theta)$, with $H(\Theta)$ the Shannon entropy of the prior, and value ε times the value of full information, so value and capacity vanish together at the same rate.

The formula also settles when the slope is infinite. That is so exactly when f is kinked at the prior, f being then of order $\|\pi - \mu\|$ where D_{KL} is of order $\|\pi - \mu\|^2$; with finitely many actions this happens when two actions with different state-contingent payoffs are both optimal at μ, which is a knife-edge in the prior. That this is the mechanism is Chade and Schlee's own observation: with more than one action optimal at the prior, a small

movement of the posterior can move the set of maximisers a long way, so that a little information can carry a positive marginal value. Their Example 7 is such a problem: with $u(a,\theta) = a\theta$ on $A = [0,1]$ and a symmetric prior on $\Theta = \{-1,1\}$ every action is optimal at μ, so $f(\pi) = (E_\pi\theta)^+$ is of order $\|\pi - \mu\|$ and the finite positive slope they report in their own index is $+\infty$ in capacity. That kink is the only source of an infinite slope, and excluding it is the only restriction we impose on preferences.

Assumption 1. $V'(0^+) < \infty$.

Assumption 1 holds in both of the problems above, and the logarithmic portfolio problem is transparent in these terms. With the bets of Section 1 at odds X_i, the portfolio that is optimal at posterior π is $a_i = \pi_i$, so $\varphi(\pi) = \sum_i \pi_i \log X_i - H(\pi)$, while the prior-optimal portfolio $a = \mu$ delivers $\ell(\pi) = \sum_i \pi_i \log X_i + \sum_i \pi_i \log \mu_i$. Hence

$$f(\pi) = \varphi(\pi) - \ell(\pi) = D_{\mathrm{KL}}(\pi\|\mu) \tag{4}$$

identically: every posterior is worth exactly its own divergence from the prior. The ratio in (3) is therefore 1 at every π rather than approached along some sequence, the gross value of any experiment is its capacity, and $V(\kappa) = \min\{\kappa, H(\Theta)\}$ with $V'(0^+) = 1$. Beyond $\kappa = H(\Theta)$ the state is known and further capacity is worthless, so V is linear on $[0, H(\Theta)]$ and flat thereafter. Equation (4) is Arrow's statement that the value of a channel is precisely its rate of transmission (Arrow, 1971, p. 112), and his theorem that logarithmic utility is the only utility for which that value is independent of the odds (Arrow, 1971, pp. 109–111) says that (4) holds under logarithmic utility and under no other. For a normal state under quadratic loss, which is the workhorse of rate distortion theory (e.g. Behringer and Belavkin, 2027, Chapter 4), with prior variance σ_0^2 and stakes $k > 0$, the optimal experiment at capacity κ is additive Gaussian with posterior variance $\sigma_0^2 e^{-2\kappa}$, so that

$$V(\kappa) = k\sigma_0^2\left(1 - e^{-2\kappa}\right), \qquad V'(0^+) = 2k\sigma_0^2, \tag{5}$$

which is what (3) returns without differentiating V: here $f(\pi) = k\left(m_\pi - m_0\right)^2$ depends on the posterior only through its mean m_π, while $D_{\mathrm{KL}}(\pi\|\mu) \geq \left(m_\pi - m_0\right)^2 / 2\sigma_0^2$ with equality for a Gaussian posterior of variance σ_0^2, so the ratio is bounded by $2k\sigma_0^2$ and attained at every such posterior.

3. The order of capacity

Lemma 2. *Let $\{S(t)\}_{t\geq 0}$ be a family of experiments in which $S(0)$ is uninformative, so that the signal density at $t=0$ is the same for every state and may be written $p_0(\cdot)$, and let $p_t(\cdot \mid \theta)$ be absolutely continuous with respect to p_0 for every t and θ, with likelihood ratio*

$$r_t(s,\theta) := \frac{p_t(s \mid \theta)}{p_0(s)} = 1 + t\, g(s,\theta) + t^2 h(s,\theta) + o(t^2), \tag{6}$$

the remainder being $o(t^2)$ in $L^1(p_0 \otimes \mu)$, with $g \in L^2(p_0 \otimes \mu)$ not degenerate in θ, the state being distributed according to the prior μ. Then

$$\kappa(t) \;=\; I\left(\Theta, S(t)\right) \;=\; \frac{J}{2}\, t^2 + o(t^2), \qquad J \;:=\; E_{p_0}\{\mathrm{Var}_\Theta\, g(s,\Theta)\} > 0, \tag{7}$$

so that $\kappa'(0) = 0$ and $\mathrm{d}t/\mathrm{d}\kappa \to \infty$ as $\kappa \downarrow 0$.

The proof is in Appendix A.

The normalisation that removes the first-order term is the one that removes the corresponding term in the proof of Radner and Stiglitz: the first-order term of a differentiable perturbation of a probability distribution integrates to zero. An index moving the likelihood at rate t therefore moves information at rate t^2.[4]

The second-order coefficient drops out because h reaches order t^2 only through the leading 1 in r_t, and so enters linearly, as its deviation from its average across states, which averages to zero; pairing it with tg would cost a further power of t. Only g survives, and only quadratically, through its dispersion across states: what the signal reveals at leading order is the extent to which the likelihood moves differently in different states, so a perturbation common to all of them, however large, shifts the marginal by the same amount and carries no information. This is what Var_Θ in (7) records, and the non-degeneracy of g in θ assumed in the lemma is the condition for $J > 0$.[5]

[4] It is also the term that vanishes in the proof of Lemma 1, where the residual posterior μ_p differs from the prior by $O(p)$ and contributes $O(p^2)$ to the capacity of τ_p.

[5] De Lara and Gilotte (2007) reach the Radner–Stiglitz conclusion by a different route, giving a sufficient condition stated on the information structure alone: vanishing of their infinitesimal information distance variation implies a zero marginal value of information at the null independently of preferences, and the condition is tight, a positive value of it yielding a payoff function whose marginal value at the null is positive. Lemma 2 works instead through the order of mutual information in the index, and in the present argument preferences re-enter through Assumption 1.

The second order in (7) is delivered by the hypothesis that the family perturbs a fixed noise distribution: the signal law stays on the support of p_0 and moves only through its density, so a small t moves every posterior only slightly, and divergence is quadratic in that movement. Not every family is of that kind, and the one behind $V'(0^+) > 0$ in Lemma 1 is not. A family that reveals θ outright with probability t and says nothing otherwise, which is the τ_p of that proof at $p = t$, puts mass on signals that p_0 excludes, so no expansion (6) exists; its capacity is $tH(\Theta) + O(t^2)$, of first order, because the posterior does not move slightly with high probability but a fixed distance with small probability.

Proposition 1. *Let $G(t) := V(\kappa(t))$ under Assumption 1 and Lemma 2. Then $G(t) = V'(0^+)\frac{J}{2}t^2 + o(t^2)$, so $G'(0) = 0$, and for any cost γ with $\gamma(0) = 0$ and $\gamma'(0) > 0$ there is $\bar{t} > 0$ such that $G(t) - \gamma(t) < 0$ on $(0, \bar{t})$. Expressed in capacity units, such a cost is $C(\kappa) = \gamma'(0)\sqrt{2\kappa/J} + o(\sqrt{\kappa})$, with $C'(0^+) = +\infty$.*

Proof. Under Assumption 1 the right derivative is finite, so $V(\kappa) = V'(0^+)\kappa + o(\kappa)$ as $\kappa \downarrow 0$.; composing with (7) and using $\kappa(t) = O(t^2)$, so that the remainder is $o(\kappa(t))$ and hence $o(t^2)$, gives G. Since G is of order t^2 and therefore negligible against t, while $\gamma(t) = \gamma'(0)t + o(t)$, the difference is $G(t) - \gamma(t) = -\gamma'(0)t + o(t)$, negative for small $t > 0$ because $\gamma'(0) > 0$. Inverting (7) as $\kappa(t) = \frac{J}{2}t^2 + o(t^2) = \frac{J}{2}t^2(1 + o(1))$ gives $t(\kappa) = \sqrt{2\kappa/J}\,(1 + o(1))$, substituted into $\gamma(t) = \gamma'(0)t + o(t)$ gives C with $C'(0^+) = +\infty$. □

The two coordinates locate the negative sign at the origin differently. In the index the cost is unremarkable, a finite $\gamma'(0) > 0$, and what leaves the first increment not worth buying is the flat gross value, $G'(0) = 0$; the cost alone would not do it. In capacity the cost alone does it, $C'(0^+) = +\infty$ against a finite $V'(0^+)$, and the gross value is no longer flat.

Arrow (1985, p. 306) accepts the Radner–Stiglitz theorem and locates the hypothesis that fails in the likelihood: along a family indexed by its rate of transmission, the conditional density has infinite derivative at the null. Proposition 1 locates the same failure in the cost, and the two statements are one change of variable. A family of experiments carries two coordinates: the index t in which the modeller happened to write it down, and the capacity κ it delivers. Lemma 2 supplies the derivative of the change of variable between them, $\mathrm{d}t/\mathrm{d}\kappa = (2J\kappa)^{-1/2}\,(1 + o(1))$, which diverges at the origin precisely because $\kappa'(0) = 0$, and by the chain rule anything expressed in

the index inherits that divergence when re-expressed in capacity. Applied to the conditional density it gives $\partial p/\partial\kappa = (\partial p/\partial t)\,\mathrm{d}t/\mathrm{d}\kappa \to \infty$, which is his statement; applied to a cost linear in t it gives $C'(\kappa) = \gamma'(0)\,\mathrm{d}t/\mathrm{d}\kappa \to \infty$, which is Proposition 1. One factor, $\mathrm{d}t/\mathrm{d}\kappa$, multiplied by $\partial p/\partial t$ in his case and by $\gamma'(0)$ in ours.

A cost linear in capacity is the mutual-information cost $\lambda\kappa$ of the Rational Inattention literature, whose marginal cost is the constant λ per nat and which Proposition 1 does not touch. A cost with strictly positive marginal cost in a smoothly resolving index charges at a positive rate for the resolving power of the instrument, whereas what the decision maker values is the information that resolving power delivers, which is nothing at first order. Acquiring the first κ nats requires an index movement of $\sqrt{2\kappa/J}$, so the average cost per nat is $\gamma'(0)\sqrt{2/(J\kappa)}$, unbounded as $\kappa \downarrow 0$ although the cost per unit of index is bounded away from zero and finite. Only the first-order behaviour of γ at the null is at work: what the schedule does at larger t is irrelevant to the divergence.

Radner and Stiglitz (1984) perform that change of variable themselves, in an example of their own. Their index of informativeness, which plays the role of our t, is in their linear prediction example the correlation coefficient: with $t = \rho$ they obtain $V = t^2 - 1 - \gamma(t)$ and $V'(0) = -\gamma'(0) < 0$, their equations (24) and (26),[6] while with $t = \rho^2$ they obtain $V = t - 1 - \widetilde{\gamma}(t)$ with $\widetilde{\gamma}(t) = \gamma(t^{1/2})$ and $V'(0) = 1 - \widetilde{\gamma}'(0)$, which may be positive, their (24′) and (25′). The two are reconciled, they observe, by the failure of their own differentiability hypothesis under the second indexing. For a bivariate normal $\kappa = -\frac{1}{2}\log(1-\rho^2) = \frac{1}{2}\rho^2 + O(\rho^4)$, so their second index is twice the capacity to leading order: the reindexing they carry out is the passage to capacity units, and $\widetilde{\gamma}'(0) = +\infty$ for a cost linear in ρ is Proposition 1 in that example.

Radner and Stiglitz treat the portfolio model themselves, and levy their fee on the index. The charge is a smooth function of it, deducted from initial wealth, so that the cost is the utility of the wealth foregone, and they obtain

[6] With a cost linear in the index their (24) reads $V(t) = t^2 - 1 - \gamma'(0)t$, which is convex, so the stationary point $t = \gamma'(0)/2$ delivered by the first-order condition is a minimum, the optimum is an endpoint of the feasible range and no interior informativeness is ever chosen; the same holds after their reindexing, where $V(t) = t - 1 - \gamma'(0)\sqrt{t}$ is again convex, so demand is bang-bang under either choice, moves discontinuously with the cost, and the discontinuity belongs to the problem rather than to the parametrisation.

$V = I(t) + f + \log[W - \gamma(t)] + \sum_s \phi_s \log \phi_s$ with $f = \sum_s \phi_s \log r_s$; since $I'(0) = 0$ they conclude $V'(0) = -\gamma'(0)/W < 0$, their equations (32), (37) and (38). In capacity that charge reads

$$C(\kappa) = \log W - \log\{W - \gamma(t(\kappa))\} = \frac{\gamma'(0)}{W}\sqrt{\frac{2\kappa}{J}} + o\left(\sqrt{\kappa}\right), \tag{8}$$

with $C'(0^+) = +\infty$: finite per unit of index, but unbounded per nat as $\kappa \downarrow 0$, which is Proposition 1 in the model.[7]

Arrow's two examples fall on opposite sides of Lemma 2. The first, the channel of the portfolio model, is a smooth perturbation of a fixed noise distribution merely reparametrised by its rate of transmission, so the lemma applies: the infinite derivative he reports is the change of variable, and it sends the cost per nat to infinity. The second fails the hypothesis outright. At zero noise precision the noise is infinitely diffuse and there is no density for the family to perturb, so no p_0 exists; capacity is then of first order, being $\kappa = \frac{1}{2}\log(1 + h_\varepsilon/h_0)$ at prior precision h_0 and noise precision h_ε, so that $\mathrm{d}\kappa/\mathrm{d}h_\varepsilon = 1/(2h_0)$ at the null and a cost linear in precision has finite marginal cost per nat. There the infinite derivative belongs to the family itself rather than to any change of variable, which is why both sides of the comparison in Section 4 are finite and his sampling optimum interior.

What the Radner–Stiglitz hypotheses restrict is therefore the family together with its parametrisation and the cost defined on it, the restriction on the decision problem reducing to the exclusion of a tie at the prior. The primitives enter only through V, which by Lemma 1 is an envelope over all experiments of a given capacity rather than a value along a given family, and so is invariant to relabelling where C is not; what this assumes in exchange is that capacity is acquired efficiently. What Radner and Stiglitz rule out is accordingly a class of cost systems for information.

[7] The model also dispenses with the composition in Proposition 1: by (4) the gross value of an experiment is its capacity, so $G(t) = \kappa(t)$ and the flat gross value at the null, $G'(0) = 0$, is exactly $\kappa'(0) = 0$ from Lemma 2, while the same value counted in capacity rises at rate $V'(0^+) = 1$.

4. When a little information pays

Lemma 1 makes the gross value concave in capacity and gives its slope at the origin; what remains is to compare that slope with the marginal cost of the first nat. Neither one-sided limit need be finite, and neither function need be differentiable, for the comparison to be available.

Proposition 2. *Let C be convex with $C(0) = 0$. If*

$$C'(0^+) < V'(0^+) \tag{9}$$

then $V(\kappa) - C(\kappa) > 0$ for all sufficiently small $\kappa > 0$ and every optimal capacity is strictly positive. If $C'(0^+) \geq V'(0^+)$ then $V(\kappa) - C(\kappa) \leq 0$ for every $\kappa \geq 0$, and zero capacity is optimal.

Proof. By Lemma 1, $V(\kappa)/\kappa$ is nonincreasing with limit $V'(0^+)$; by convexity, $C(\kappa)/\kappa$ is nondecreasing with limit $C'(0^+)$. Under (9) the first exceeds the second for small κ; otherwise $C(\kappa)/\kappa \geq C'(0^+) \geq V'(0^+) \geq V(\kappa)/\kappa$ for every κ. □

The proof takes no derivatives and does not use Assumption 1. Concavity of V with $V(0) = 0$ and convexity of C with $C(0) = 0$ make the two difference quotients monotone, so their limits exist in $[0, +\infty]$ without further hypothesis and the comparison is between the quotients themselves. The criterion therefore returns a verdict even in the cases where Radner and Stiglitz (1984) cannot form their derivative at all, $V'(0^+) = +\infty$ included.

Assumption 1 is what gives the criterion something to decide. When it fails, $V'(0^+) = +\infty$ and (9) holds against every charge whose marginal cost at zero is finite, so a little information always pays and the comparison has no work to do. Under it both sides are finite numbers, and which cost systems satisfy (9) becomes a substantive question, answered for the leading specifications by Corollary 1.

Corollary 1. *A cost differentiable at the null with strictly positive slope in the index of a family satisfying the hypotheses of Lemma 2 has $C'(0^+) = +\infty$ by Proposition 1 and so fails* (9)*. Being of order $\sqrt{\kappa}$ at the origin it is not convex there, so it falls outside the hypotheses of Proposition 2, whose second clause does not apply: the optimum is at a corner of the feasible range rather than at zero, as in footnote 6. The mutual-information cost $C(\kappa) = \lambda\kappa$ of the Rational Inattention literature (Maćkowiak et al., 2023) has $C'(0^+) = \lambda$. The*

log-likelihood-ratio cost of Pomatto et al. (2023) charges a normal experiment in proportion to its precision, so a measurement of a state with prior precision h_0 through noise of variance σ_ε^2 costs $b\sigma_\varepsilon^{-2}$; since $\sigma_\varepsilon^{-2} = h_0\left(e^{2\kappa}-1\right)$, this reads $C(\kappa) = bh_0\left(e^{2\kappa}-1\right)$ with $C'(0^+) = 2bh_0$. [8] *Both are finite, and each satisfies* (9) *precisely when its $C'(0^+)$ falls below $V'(0^+)$.*

In Arrow's portfolio problem both sides of (9) can be computed. With logarithmic utility and a constant cost per unit of channel capacity, he observes, the value of information and its cost are both proportional to the amount of information, so the net value is $\left(V'(0^+) - C'(0^+)\right)\kappa$ and demand is all, nothing, or anything according to the sign of that difference, the last requiring the two slopes to coincide exactly (Arrow, 1971, p. 112).

Finding this unreasonable, he diagnoses the difficulty as dimensional, cost being measured in resources and value in utility, and charges the cost against initial resources rather than against the outcome (Arrow, 1971, p. 113). In the notation used here that choice makes the utility cost $C(\kappa) = \ln A - \ln\left(A - c\kappa\right)$, strictly convex with $C'(0^+) = c/A$.[9] Against $V'(0^+) = 1$, (9) then reads $c < A$, his condition for a strictly positive purchase (Arrow, 1985, p. 306); the optimum itself follows from $C'(\kappa^*) = V'(\kappa^*) = 1$, that is $c/\left(A - c\kappa^*\right) = 1$, which is his $R^* = (A-c)/c$, and with resources normalised to one it is his $R = -1 + (1/C)$ (Arrow, 1971, p. 114), his C being the unit cost written c here, valid where it falls below $H(\Theta)$, the corner at full information being the optimum otherwise. The investor, the securities and the prior are those of (8), and so is the gross value, so what separates a marginal cost of c/A from one of $+\infty$ is nothing but whether the fee is levied on the index or on the capacity it delivers.

Proposition 2 is the general form of the comparison Arrow made twice,

[8] Their axioms deliver, for finitely many states, a cost that is a weighted sum of divergences between the state-conditional signal distributions, with the weights fixed by the axioms. For a normal state that theorem does not apply, so what is taken over here is the functional form, the constant $b > 0$ multiplying it being left free rather than determined by their axioms. Positivity is all the corollary uses, since it needs only $C'(0^+) = 2bh_0$ to be finite and non-zero.

[9] Under logarithmic utility the optimal bets are $a_i = \pi_i$ whatever the wealth, so a fee $c\kappa$ paid out of A leaves expected utility $\ln\left(A - c\kappa\right) + E\left\{\ln r_x\right\} - H(\Theta) + \kappa$, against $\ln A + E\left\{\ln r_x\right\} - H(\Theta)$ at $\kappa = 0$. The difference is $\kappa - [\ln A - \ln\left(A - c\kappa\right)]$, so C is the utility of the wealth foregone, with $C'(\kappa) = c/\left(A - c\kappa\right)$ and $C''(\kappa) = c^2/\left(A - c\kappa\right)^2 > 0$: the logarithm turns a fee linear in κ into a strictly convex utility cost of finite slope at the origin.

once for each of his examples, and his two costs are the two the later literature took up, the constant rate per unit of transmission by Maćkowiak et al. (2023) and the charge proportional to sample size by Pomatto et al. (2023), who open by quoting Arrow on the need to formulate reasonable cost functions for information structures (Pomatto et al., 2023, n. 1).

The capacity constraint of Sims (2003), under which any experiment below a bound is free, is the limiting case. Read as a cost it is zero up to the bound and infinite beyond, which is convex with $C(0) = 0$, so Proposition 2 applies to it and the constraint binds at $\kappa^* = \bar{\kappa}$. What is primitive there is the bound and not a cost, and the shadow-price reading keeps it so: the price per nat is derived from the bound, the multiplier $\lambda^*(\bar{\kappa}) \in \partial V(\bar{\kappa})$ that the constraint generates (Rockafellar, 1974). That schedule is the inverse demand for capacity, decreasing because V is concave, so its intercept $V'(0^+)$ is the choke price: the cost per nat at and above which nothing is bought, which is the second part of Proposition 2 read as a demand curve.[10]

The constraint meets (9) on either reading, though only one of the two is a cost of the kind Radner and Stiglitz (1984) assume. Read as a cost schedule the constraint has zero marginal cost near the origin, so (9) holds by construction and their hypothesis of a strictly positive marginal cost is not met at all. Read as the price it generates the marginal cost is the strictly positive λ^*, and there concavity does the work twice. It first delivers the multiplier: $\partial V(\bar{\kappa})$ is non-empty at every interior $\bar{\kappa}$, and $\lambda^* \in \partial V(\bar{\kappa})$ says exactly that $V(\kappa) - \lambda^*\kappa \leq V(\bar{\kappa}) - \lambda^*\bar{\kappa}$ for every κ, so the bound and the linear cost $C(\kappa) = \lambda^*\kappa$ select the same experiment, the two being representations of one choice. That cost has $C(0) = 0$ and $C'(0^+) = \lambda^*$. Concavity then orders the subdifferentials, so $\lambda^* \leq V'(0^+)$. The two steps together give $C'(0^+) < V'(0^+)$, which is (9) for the price the constraint actually generates.

That $\bar{\kappa}$ is bought is assumed rather than concluded, the constraint being taken to bind; what the argument adds is that the price supporting it never reaches the choke price, so the configuration Radner and Stiglitz (1984) describe, a strictly positive marginal cost at which the first increment does not pay, is not the shadow price of any capacity constraint, whatever the primitives of the problem.[11] Where V is affine the inequality becomes an

[10] The Stratonovich/Shannon Value of Information framework (Stratonovich, 1975/2020) underlying these results is developed in Behringer and Belavkin (2027), and the endogenous price generated by a capacity bound in Behringer (2026a, Thms. 1–3).

[11] The multiplier is recovered as $\lambda^*(\bar{\kappa}) \in \partial V(\bar{\kappa})$ by convex duality, without differentia-

equality[12] and Arrow's indifference among every capacity returns. That the two readings agree at all is a consequence of counting in capacity: in the index the composed value G of Proposition 1 is convex near the origin, so bounding the index and pricing it select different experiments, the bound in the one case and a corner in the other.

5. What changes

Neither implication drawn by Radner and Stiglitz (1984) survives the passage to capacity units, in which V is concave by Lemma 1 and C is convex. Condition (9) enters the first of them and not the second.

First, increasing returns. By Lemma 1 the gross value is concave in capacity throughout, and under (9) the net value is increasing from the origin. Increasing returns are present in the smooth index t, because κ is convex in t there, and absent in κ. The nonconcavity is therefore a property of the parametrisation rather than of the information technology, and the generality of the nonconcavity in Chade and Schlee (2002, §5) is generality across families: their value function runs along a given family, where (2) is the upper envelope over all experiments of a given capacity, and a family need not deliver, at each capacity, the most valuable experiment of that capacity.

Second, discontinuous demand. Under a convex cost the net value is concave, so at every price the set of optimal capacities is an interval and demand is a monotone correspondence: a smooth decreasing function where V is strictly concave and differentiable, constant over an interval of prices at a kink of V, and discontinuous only where V is affine. With the linear cost $\lambda\kappa$ the first-order condition is $V'(\kappa^*) = \lambda$, the marginal value of capacity equated to its price; under (5) it reads $2k\sigma_0^2 e^{-2\kappa^*} = \lambda$ and gives $\kappa^* = \frac{1}{2}\log(2k\sigma_0^2/\lambda)$, which falls continuously to zero as λ rises to the choke price $V'(0^+)$ and is zero above it.

Their jump survives reindexing, because a reindexing carries the cost with it: linear in t becomes $C(\kappa) = \gamma'(0)\sqrt{2\kappa/J} + o(\sqrt{\kappa})$ in capacity, concave at

bility of V or strict concavity. Both fail in the logarithmic portfolio model above, where V is affine up to $H(\Theta)$ and kinked there, so $\partial V(\bar{\kappa})$ is a set rather than a number; nothing in the comparison depends on which of its elements is taken, concavity bounding every one of them by $V'(0^+)$.

[12] With V affine of slope s the difference quotients $V(\kappa)/\kappa$ no longer decrease, so $V'(0^+) = s = \lambda^*$, and the representing cost being $\lambda^*\kappa$ gives $C'(0^+) = V'(0^+)$.

the origin, and a cost concave there against a concave value leaves the net value nonconcave and the optimum at a corner. That a change of variable cannot by itself remove the jump is Chade and Schlee (2002, Cor. 4), of which footnote 6 is an instance. What Proposition 2 changes is the cost and not the index: a charge linear in capacity, or linear in precision, is not the transform of any charge with strictly positive marginal cost in a smoothly resolving index, and their result does not reach it. What removes the jump is therefore not the change of variable but the convexity of the cost in capacity assumed in Proposition 2, which makes the net value concave and leaves (9) to decide whether the optimum is zero or strictly positive. Linearity in capacity is sufficient for that convexity and not necessary: a cost linear in precision is convex in capacity too, which is why Arrow's sampling optimum is interior.

Acknowledgement

I thank Roman V. Belavkin for early comments on the topic.

Appendix A. Proofs of the Lemmata

Proof of Lemma 1. Let τ_1, τ_2 be feasible at κ_1, κ_2 and let Z be a public randomisation, independent of θ, equal to 1 with probability α. Observing Z and then the selected signal is itself an experiment, whose capacity is $\alpha\kappa_1 + (1-\alpha)\kappa_2$ and whose value is the corresponding mixture of the two values, so the set of attainable capacity-value pairs is convex and V is concave; $V(0) = 0$ and monotonicity are immediate. Concavity with $V(0) = 0$ makes the slope $V(\kappa)/\kappa$ nonincreasing, since $V(\kappa_1) \geq (\kappa_1/\kappa_2)\, V(\kappa_2)$ for $0 < \kappa_1 < \kappa_2$, so its limit at the origin is its supremum, which is the first equality in (3).

For the second, write η for the supremum on the right of (3). Any feasible τ has $E_\tau f \leq \eta\, E_\tau D_{\mathrm{KL}}(\pi\|\mu) \leq \eta\kappa$, so $V'(0^+) \leq \eta$.

Conversely fix π with bounded density with respect to μ and put $\mu_p := (\mu - p\pi)/(1-p)$ for p small enough that $p\,\mathrm{d}\pi/\mathrm{d}\mu \leq 1$, so that μ_p is a probability measure and $\tau_p := p\delta_\pi + (1-p)\delta_{\mu_p}$ is Bayes plausible, has capacity $p\, D_{\mathrm{KL}}(\pi\|\mu) + (1-p) D_{\mathrm{KL}}(\mu_p\|\mu) = p\, D_{\mathrm{KL}}(\pi\|\mu) + O(p^2)$, since μ_p differs from μ by $O(p)$ and divergence from a fixed measure vanishes to second order in that difference, and has value at least $p\, f(\pi)$, since $f \geq 0$. Letting $p \downarrow 0$ gives

$$V'(0^+) \;\geq\; \frac{f(\pi)}{D_{\mathrm{KL}}(\pi\|\mu)}.$$

Taking the supremum over such π yields $V'(0^+) \geq \eta$: the restriction to bounded densities does not lower the supremum, since only π with $D_{\mathrm{KL}}(\pi\|\mu) < \infty$ can matter, f being bounded, and for such π the truncations π_n with density proportional to $\min\{\mathrm{d}\pi/\mathrm{d}\mu, n\}$ converge to π in total variation, so that $f(\pi_n) \to f(\pi)$, f being Lipschitz in total variation with constant $\sup|u|$, while $D_{\mathrm{KL}}(\pi_n\|\mu) \to D_{\mathrm{KL}}(\pi\|\mu)$ by monotone convergence. Together with $V'(0^+) \leq \eta$ established above this gives (3). □

Proof of Lemma 2. Since each $p_t(\cdot \mid \theta)$ integrates to one for every t, integrating (6) against p_0 and matching powers of t gives $\int p_0 g(\cdot,\theta) = \int p_0 h(\cdot,\theta) = 0$ for every θ. Writing $\bar{g} := E_\Theta g$ and $\bar{h} := E_\Theta h$, the marginal density is $p_t(s) = p_0(s)\left[1 + t\bar{g} + t^2\bar{h} + o(t^2)\right]$, so that with $\bar{r}_t := p_t(s)/p_0(s)$,

$$\log \frac{r_t}{\bar{r}_t} = t\,(g - \bar{g}) + t^2\left[\left(h - \bar{h}\right) - \bar{g}\,(g - \bar{g}) - \tfrac{1}{2}\,(g - \bar{g})^2\right] + o(t^2). \qquad \text{(A.1)}$$

Since $p_t(s \mid \theta)\,\mathrm{d}s = p_0(s) r_t\,\mathrm{d}s$ and $p_t(s \mid \theta)/p_t(s) = r_t/\bar{r}_t$, the mutual information is

$$\kappa(t) \;=\; E_\Theta \int p_0\, r_t \log \frac{r_t}{\bar{r}_t}\,\mathrm{d}s.$$

Multiplying (A.1) by $r_t = 1 + tg + O(t^2)$ gives

$$r_t \log \frac{r_t}{\bar{r}_t} = t\,(g - \bar{g}) + t^2\left[\left(h - \bar{h}\right) + (g - \bar{g})^2 - \tfrac{1}{2}\,(g - \bar{g})^2\right] + o(t^2), \qquad \text{(A.2)}$$

the square arising because the product of tg in r_t with $t\,(g - \bar{g})$ in the logarithm contributes $g\,(g - \bar{g})$, against $-\bar{g}\,(g - \bar{g})$ from the bracket of (A.1), and $g\,(g - \bar{g}) - \bar{g}\,(g - \bar{g}) = (g - \bar{g})^2$. Taking $E_\Theta \int p_0(\cdot)\,\mathrm{d}s$ of (A.2) term by term, the term of order t vanishes because $E_\Theta\{g - \bar{g}\} = 0$ at every s, by definition of $\bar{g}$; at order t^2 the same argument removes the contribution in h, since $E_\Theta\left\{h - \bar{h}\right\} = 0$, and what survives is

$$E_{p_0}\left\{E_\Theta\,(g - \bar{g})^2 - \tfrac{1}{2} E_\Theta\,(g - \bar{g})^2\right\} = \tfrac{1}{2} E_{p_0}\{\mathrm{Var}_\Theta\, g\} = \frac{J}{2}.$$

The expansion $\kappa(t) = \frac{J}{2}t^2 + o(t^2)$ implies $\kappa'(0) = 0$, so that

$$\frac{\mathrm{d}t}{\mathrm{d}\kappa} = (Jt)^{-1}\,(1 + o(1)) = (2J\kappa)^{-1/2}\,(1 + o(1)) \longrightarrow \infty$$

as $\kappa \downarrow 0$. □

References

Arrow, K. J., 1971. The value of and demand for information. In: McGuire, C. B., Radner, R. (Eds.), *Decision and Organization*. North-Holland, Amsterdam, ch. 6, pp. 131–139. Reprinted in: *The Economics of Information*, Collected Papers of Kenneth J. Arrow, vol. 4, Harvard University Press, Cambridge MA, 1984, pp. 106–114. Page references are to the reprint.

Arrow, K. J., 1985. Informational structure of the firm. *American Economic Review* 75 (2), 303–307.

Behringer, S., 2026a. Hard constraints, convex duality, and the endogenous cost of information. Mimeo, Universität Bielefeld.

Behringer, S., Belavkin, R. V., 2027. *The Value of Information in Economics and Decision Theory*. Springer, forthcoming.

Chade, H., Schlee, E. E., 2002. Another look at the Radner–Stiglitz nonconcavity in the value of information. *Journal of Economic Theory* 107 (2), 421–452.

De Lara, M., Gilotte, L., 2007. A tight sufficient condition for Radner–Stiglitz nonconcavity in the value of information. *Journal of Economic Theory* 137 (1), 696–708.

Denti, T., 2022. Posterior separable cost of information. *American Economic Review* 112 (10), 3215–3259.

Kamenica, E., Gentzkow, M., 2011. Bayesian persuasion. *American Economic Review* 101 (6), 2590–2615.

Matějka, F., McKay, A., 2015. Rational inattention to discrete choices: a new foundation for the multinomial logit model. *American Economic Review* 105 (1), 272–298.

Maćkowiak, B., Matějka, F., Wiederholt, M., 2023. Rational inattention: a review. *Journal of Economic Literature* 61 (1), 226–273.

Pomatto, L., Strack, P., Tamuz, O., 2023. The cost of information: the case of constant marginal costs. *American Economic Review* 113 (5), 1360–1393.

Radner, R., Stiglitz, J. E., 1984. A nonconcavity in the value of information. In: Boyer, M., Kihlstrom, R. E. (Eds.), *Bayesian Models in Economic Theory*. Elsevier, Amsterdam, pp. 33–52.

Rockafellar, R. T., 1974. *Conjugate Duality and Optimization*. CBMS-NSF Regional Conference Series in Applied Mathematics 16. SIAM, Philadelphia.

Sims, C. A., 2003. Implications of rational inattention. *Journal of Monetary Economics* 50 (3), 665–690.

Stratonovich, R. L., 1975/2020. *Theory of Information and Its Value*. Belavkin, R. V., Pardalos, P. M., Principe, J. C. (Eds.). Springer, Cham.

Whitmeyer, M., 2025. A concavity in the value of information. *Games and Economic Behavior* 154, 200–207.