

Experimental cost of information

Tommaso Denti Massimo Marinacci Aldo Rustichini*

May 2022

Abstract

We relate two main representations of the cost of acquiring information: a cost that depends on the experiment performed, as in statistical decision theory, and a cost that depends on the distribution of posterior beliefs, as in applications of rational inattention. We show that in many cases of interest, posterior-based costs are inconsistent with a primitive model of costly experimentation. The inconsistency is at the core of known limits to the application of rational inattention in games and, more broadly, equilibrium analyses where beliefs are endogenous; we show that an experiment-based approach helps to understand and overcome these difficulties.

For Bayesian decision makers, the acquisition of information admits two standard representations: a *statistical experiment* $P : \Theta \rightarrow \Delta(X)$ mapping states of nature into probability distributions over signals, and a *random posterior* $\mu \in \Delta(\Delta(\Theta))$ detailing a probability distribution over posterior beliefs. The relationship between these two representations is well understood: given a prior belief $\pi \in \Delta(\Theta)$, every experiment P induces via Bayesian updating a random posterior $\mu = B(\pi, P)$ that satisfies the martingale property $\int_{\Delta(\Theta)} p \, d\mu(p) = \pi$; conversely, if the set of feasible experiments \mathcal{E} is rich enough, then every random posterior can be induced in this way by some experiment.¹

Analogously, we can distinguish between two main representations for the cost of acquiring information. In one representation, the cost of acquiring information depends on the experiment that the decision maker performs, and is represented by a function $h : \mathcal{E} \rightarrow \mathbb{R}_+$. This has been the standard representation for the cost of information in statistical decision theory since Wald (1950).

*Tommaso Denti: Cornell University, tjd237@cornell.edu. Massimo Marinacci: Università Bocconi and IGIER, massimo.marinacci@unibocconi.it. Aldo Rustichini: University of Minnesota, aldo.rustichini@gmail.com. We thank the editor, Sylvain Chassang, and the referees for their comments and suggestions. We also thank Francesco Fabbri, Fabio Maccheroni, Jeffrey Mensch, Stephen Morris, and Alex Wolitzky for useful discussions. Massimo Marinacci gratefully acknowledges the financial support of ERC (grant INDIMACRO). Aldo Rustichini thanks the U.S. Department of Defense (contract W911NF2010242).

¹We formally define these objects in Section 1.

A different perspective has emerged from the work of Sims (2003) on *rational inattention*. In many applications of rational inattention, the cost of acquiring information depends on the decision maker’s random posterior, and is represented by a function $c : \Delta(\Delta(\Theta)) \rightarrow \mathbb{R}_+$. In a widely adopted specification (e.g., Matějka and McKay, 2015), the quantity $c(\mu)$ is the expected reduction in the entropy of beliefs.²

In this paper we study the relationship between these two representations for the cost of information $h : \mathcal{E} \rightarrow \mathbb{R}_+$ and $c : \Delta(\Delta(\Theta)) \rightarrow \mathbb{R}_+$. We show that in many cases of interest, posterior-based costs are *inconsistent* with a primitive model of costly experimentation.

A cost on experiments h obviously generates a cost on random posteriors c_h :

$$c_h(\mu) = \inf\{h(P) : B(\bar{\mu}, P) = \mu\} \tag{1}$$

where $\bar{\mu} = \int_{\Delta(\Theta)} p \, d\mu(p)$. For a decision maker with prior $\bar{\mu}$, the quantity $c_h(\mu)$ represents the least expensive way to obtain the random posterior μ by performing an experiment P whose cost is $h(P)$. For example, $h(P)$ could be the market price of the information represented by P ; $h(P)$ could also incorporate non-pecuniary factors that limit the acquisition of information, such as time, effort, and cognitive resources.

We show that for many costs on random posteriors c used in rational inattention—such as the entropy-based cost—there is *no* cost on experiments h that generates them; that is, there is no h such that $c = c_h$. We also consider the more general case in which the decision maker can perform multiple experiments in sequence: the inconsistency result extends to it.

The core issue is the relationship between the cost of information and prior beliefs. Intuitively, for a generic cost on random posteriors c , the quantity $c(B(\pi, P))$ may arbitrarily depend on π . On the other hand, $c_h(B(\pi, P))$ depends on π in a structured way, as detailed by (1). Such a structure imposes discipline on the relationship between the cost of information and prior beliefs. For example, we show that properties of $c(B(\pi, P))$ that appear technical at first glance—e.g., continuity across π —are instead substantial, while other properties that may seem strong—e.g. concavity in π —are in fact natural.

The inconsistency of posterior-based costs with a primitive model of costly experimentation becomes concerning when prior beliefs are endogenous, for example, in applications to games. As shown by Ravid (2020), games with rationally inattentive players present a number of difficulties. In the game he proposes, a seller makes a take-it-or-leave-it offer to a buyer for the purchase of an indivisible good; the buyer has to exert costly effort to learn the details of the seller’s offer—think of the offer as a complex contract with many add-ons. The buyer’s “prior”—that is, the buyer’s belief before acquiring information about the offer—is her conjecture about the seller’s strategy, which is endogenous to the game.

²See, e.g., Maćkowiak, Matějka and Wiederholt (2021) for a review of the literature.

Ravid assumes that the buyer’s cost of information is the expected reduction in the entropy of beliefs. He finds a multiplicity of counterintuitive equilibria where the seller’s offer is deterministic and the buyer perfectly monitors deviations at no cost; in particular, he shows that any division of the surplus can arise in equilibrium. These equilibria crucially rely on a particular form of dependence of the cost of information on prior beliefs: when the prior is degenerate, the expected reduction in the entropy of beliefs is always zero; hence, all experiments are costless, even experiments that allow perfect monitoring.

To overcome the multiplicity issue, Ravid proposes a refinement in the spirit of Selten’s perfect equilibrium. The refinement may work for the entropy-based cost but it may not work for more general posterior-based costs.

We revisit the ultimatum game of Ravid (2020) by taking a cost on experiments as primitive. We make the basic assumption that every experiment that carries some information has a positive cost, possibly small. We show that, as a result, no equilibrium refinement is needed and the analyst can obtain sharp predictions with transparent arguments and without functional-form assumptions on the cost of information.

We agree with Ravid that the multiplicity of equilibria generated by posterior-based costs is not compelling. Ravid argues that the solution concept is problematic and proposes a refinement. We believe instead that the problem is with the primitives of the game, in particular with the specification of the cost of information. The experiment-based approach we propose also has practical advantages, for it does not rely on delicate perturbations of equilibrium strategies. Our analysis extends to general games with costly monitoring.

Related literature. Since Sims (2003), most research on information acquisition has taken a cost on random posteriors as primitive. Among the few exceptions are Mensch (2018), Pomatto, Strack and Tamuz (2020), and Morris and Strack (2019).

A classic question in statistical decision theory is how to complete the Blackwell order, which is a partial order on experiments. Mensch (2018) proposes a completion that satisfies an independence condition in the spirit of the von Neumann-Morgenstern axiom in risk theory. Mensch provides several alternative representations of this completion. In one representation, P dominates Q if and only if $c(B(\pi^*, P)) \geq c(B(\pi^*, Q))$, where c is a cost on random posteriors and π^* is a fixed reference prior.

Fixing a reference prior is a convenient way to generate costs on experiments starting from costs on random posteriors. In the paper we apply such normalization to the entropy cost of Matějka and McKay (2015) and obtain an experimental version of their logit representation of optimal information acquisition, which is a cornerstone for applications of rational inattention.

To model information production, Pomatto, Strack and Tamuz (2020) provide an ax-

onomic foundation to a cost on experiments given by

$$h(P) = \sum_{\theta, \tau \in \Theta} \beta(\theta, \tau) D_{KL}(P_\theta \| P_\tau)$$

where $\beta(\theta, \tau)$ is a non-negative number and $D_{KL}(P_\theta \| P_\tau)$ is the Kullback-Leibler divergence of P_τ from P_θ . To relate their work and rational inattention, they compute the induced cost on random posteriors— c_h in the language of our paper. In the paper we discuss alternative functional-forms for costs on experiments that can be useful in applications.

Morris and Strack (2019) study a sequential sampling problem in the spirit of Wald (1950). A decision maker observes the evolution of a Brownian motion whose drift depends on the state. The sampling cost is a function of the passage of time, which is of course independent of the decision maker’s beliefs. For a binary state, Morris and Strack characterize the induced cost on random posteriors: for $\Theta = \{\theta_0, \theta_1\}$ and $\pi(\theta_0) \in (0, 1)$,

$$c(B(\pi, P)) = \pi(\theta_0) D_{KL}(P_{\theta_0} \| P_{\theta_1}) + \pi(\theta_1) D_{KL}(P_{\theta_1} \| P_{\theta_0}).$$

In the paper we study a general version of the sequential sampling problem in Morris and Strack (2019) and derive properties of the induced cost on random posteriors. See Appendix C for a detailed discussion of Morris and Strack (2019).

The inconsistency of posterior-based costs with a primitive model of costly experimentation has been discussed several times in the literature, mostly informally and in specific contexts (e.g., Gentzkow and Kamenica, 2014, Mensch, 2018, Nimark and Sundaresan, 2019). Besides organizing and reviewing the two approaches and their connections, we believe our work makes a contribution along several dimensions.

First, the discussion in the literature has focused on whether the cost of information should depend on the decision maker’s prior beliefs. A take-away message of our paper is that prior dependence *per se* is not an issue; for example, if information acquisition is sequential, then the cost of information can exhibit a non-trivial dependence on prior beliefs—see Morris and Strack (2019) above. The issue we highlight is that such dependence, when not disciplined by a primitive model of costly experimentation, may lead us astray.

For example, a few recent papers have explored the relationship between posterior-based costs and a model of sequential learning (e.g., Bloedel and Zhong, 2021; Hébert and Woodford, 2021). In contrast with our approach, these papers allow the flow cost of information—what would be the passage of time in Morris and Strack (2019) and the function h in our paper—to depend arbitrarily on the decision maker’s evolving beliefs. As a result, they are able to generate a richer class of costs on random posteriors—for example, the entropy-based cost. Our analysis adds a caveat to these results: it is crucial that the flow cost depends

arbitrarily on the decision maker’s evolving beliefs.

We believe that the restriction we impose on the flow cost is substantive. From the perspective of rational inattention, a motivation for studying sequential information acquisition is to explain the dependence of the cost of information on the decision maker’s prior beliefs. That is a fascinating line of research. However, if the flow cost arbitrarily depends on the decision maker’s evolving beliefs, the exercise loses some of its appeal. In a circular fashion, the problem of explaining the dependence of the expected cost of information on the decision maker’s prior beliefs becomes the problem of explaining the dependence of the flow cost of information on the decision maker’s evolving beliefs. See Appendix D for a detailed discussion.

Another contribution of our work is to show that the inconsistency of posterior-based costs with a primitive model of costly experimentation is not just a theoretical concern but has practical implications, in particular, for the study of information acquisition in games. There are settings in which the inconsistency is negligible—e.g., single-agent information acquisition problems where the prior is exogenous. The inconsistency becomes salient when prior beliefs are endogenous, such as in games where players acquire information about opponents’ past actions; see also the moral hazard problem with endogenous monitoring in Mensch (2018). When prior beliefs are endogenous, our analysis highlights the conceptual and practical advantages of an experiment-based approach.

1 Setup

We review a few preliminary notions, most of them based on Bohnenblust, Shapley and Sherman (1949) and Blackwell (1951).³

1.1 Beliefs

We consider a finite set Θ of *states of nature* with typical elements θ and τ . Let $\Delta := \Delta(\Theta)$ be the set of probabilities on Θ . Depending on the context, elements of Δ will be interpreted as *prior beliefs*, generically denoted by π and ρ , or *posterior beliefs*, generically denoted by p and q . We use the symbol δ_θ for the Dirac measure concentrated on θ . We denote by Δ_+ the set of probabilities on Θ with full support.

Let $\Delta^2 := \Delta(\Delta)$ be the set of Borel probabilities on Δ . Its elements, generically denoted by μ and ν , will be interpreted as *random posteriors*, that is, probability distributions over posterior beliefs. We use the symbol δ_π for the Dirac measure concentrated on π , which

³An early monograph covering statistical experiments is Blackwell and Girshick (1954); a more recent one is Torgersen (1991). Le Cam (1996) authoritatively reviews the topic. Gollier (2001, Part VIII) provides a textbook treatment.

corresponds to no information. Full information is represented by the random posterior $\sum_{\theta} \pi(\theta) \delta_{\delta_{\theta}}$ where $\delta_{\delta_{\theta}}$ is the Dirac measure concentrated on the Dirac measure concentrated on θ . For short, we write δ_{θ}^2 instead of $\delta_{\delta_{\theta}}$.

We endow Δ^2 with the weak* topology: a sequence of random posteriors (μ_n) converges to a random posterior μ if for every continuous function $\phi : \Delta \rightarrow \mathbb{R}$, $\int \phi d\mu_n \rightarrow \int \phi d\mu$.

The probability $\bar{\mu} := \int_{\Delta} p d\mu(p)$ is the *barycenter* of μ ; we interpret it as the prior from which μ is obtained via Bayesian updating. Let Δ_{π}^2 be the set of random posteriors with barycenter π .

Elements of Δ^2 are ranked by the convex order. Let $Cv(\Delta)$ be the set of functions $\phi : \Delta \rightarrow \mathbb{R}$ that are continuous and convex.

Definition 1. The *convex order* \succeq_{cv} is a binary relation on Δ^2 defined by $\mu \succeq_{cv} \nu$ if for all $\phi \in Cv(\Delta)$, $\int \phi d\mu \geq \int \phi d\nu$.

The convex order is reflexive and transitive. It is also antisymmetric (so, a partial order): $\mu \sim_{cv} \nu$ implies $\mu = \nu$. Only random posteriors with the same barycenter can be ranked: if $\mu \succeq_{cv} \nu$ then $\bar{\mu} = \bar{\nu}$.

1.2 Experiments

We fix a Polish space X of possible *signals* with Borel σ -algebra \mathcal{X} . Let $\Delta(X)$ be the set of Borel probabilities on X , generically denoted by ξ . We endow $\Delta(X)$ with the weak* topology: a sequence of probabilities (ξ_n) converges to a probability ξ if for every bounded continuous function $\varphi : X \rightarrow \mathbb{R}$, $\int \varphi d\xi_n \rightarrow \int \varphi d\xi$.

A *statistical experiment* is a map from states into probabilities on signals, that is, from Θ into $\Delta(X)$. Typical experiments are denoted by P and Q , with $P_{\theta}(A)$ and $Q_{\theta}(A)$ the probability of event $A \in \mathcal{X}$ in state θ , respectively. Let \mathcal{E} be the set of all experiments.

The set of experiments \mathcal{E} can be regarded as the Cartesian product of copies of $\Delta(X)$, so it can be endowed with a statewise mixture operation and a product topology. For all $P, Q \in \mathcal{E}$ and $\alpha \in [0, 1]$, we define the experiment $\alpha P + (1 - \alpha)Q$ by

$$(\alpha P + (1 - \alpha)Q)_{\theta} = \alpha P_{\theta} + (1 - \alpha)Q_{\theta}.$$

A sequence of experiments (P_n) converges to an experiment P if for every state θ , the sequence of probabilities $(P_{n,\theta})$ converges to the probability P_{θ} in the weak* topology.⁴

We assume that X is rich enough to embed Δ . Formally, we assume that Δ is homeomorphic to a compact subset of X . For example, X could be a Euclidean space of dimension greater than the cardinality of Θ .

⁴There are other useful mixture operations and topologies that one can impose on experiments (see, e.g., Torgersen, 1991), but we will not need them in this paper.

Experiments are ranked via the Blackwell order. A *stochastic kernel* K is a map from $X \times \mathcal{X}$ into $[0, 1]$ such that, for every $x \in X$ and $A \in \mathcal{X}$, the set function $K(x, \cdot)$ is a probability measure and the real-valued map $K(\cdot, A)$ is measurable. For $\xi \in \Delta(X)$, we denote by $K\xi$ the probability measure on \mathcal{X} defined by

$$K\xi(A) = \int_X K(x, A) d\xi(x).$$

Definition 2. The *Blackwell order* \succeq_b is a binary relation on \mathcal{E} defined by $P \succeq_b Q$ if there exists a stochastic kernel $K : X \rightarrow \Delta(X)$ such that for all $\theta \in \Theta$, $Q_\theta = KP_\theta$.

The Blackwell order is reflexive and transitive (so, a preorder). An experiment P is *uninformative* if for all $\theta, \tau \in \Theta$, $P_\theta = P_\tau$. An experiment P is uninformative if and only if for all $Q \in \mathcal{E}$, $Q \succeq_b P$.

1.3 A Bayes map

Given a prior belief π , every experiment P induces via Bayesian updating a random posterior, denoted by $B(\pi, P)$. In this way, we define a *Bayes map*

$$B : \Delta \times \mathcal{E} \rightarrow \Delta^2$$

from pairs of priors and experiments into random posteriors.

Specifically, let $P_\pi \in \Delta(X)$ be the *predictive* probability

$$P_\pi = \sum_{\theta} \pi(\theta) P_\theta$$

that gives the ex-ante likelihood of different signal realizations. The process of Bayesian updating defines, for P_π -almost all x , a posterior belief $p_x \in \Delta$. If $\pi(\theta) > 0$, then

$$p_x(\theta) = \pi(\theta) \frac{dP_\theta(x)}{dP_\pi(x)}$$

where dP_θ/dP_π is the density of P_θ with respect to P_π .⁵ If $\pi(\theta) = 0$, then $p_x(\theta) = 0$.

When P_π has finite support, p_x follows the usual Bayes rule: for all x with $P_\pi(x) > 0$,

$$p_x(\theta) = \pi(\theta) \frac{P_\theta(x)}{P_\pi(x)}.$$

The density dP_θ/dP_π allows us to extend Bayes' rule beyond discrete experiments.

⁵If $\pi(\theta) > 0$, then P_θ is absolutely continuous with respect to P_π ; thus, by the Radon-Nikodym theorem, P_θ has a density with respect to P_π , which is unique P_π -almost surely.

The random posterior $B(\pi, P)$ is the pushforward of P_π under the function $x \mapsto p_x$. If P_π has finite support, then

$$B(\pi, P) = \sum_x P_\pi(x) \delta_{p_x}$$

where $\delta_{p_x} \in \Delta^2$ is the Dirac measure concentrated on p_x .

By the so-called ‘‘martingale property’’ of Bayesian updating, if $\mu = B(\pi, P)$, then $\bar{\mu} = \pi$. There is therefore a well-defined sense in which experiments generate random posteriors:

Definition 3. An experiment $P \in \mathcal{E}$ generates a random posterior $\mu \in \Delta^2$ if $\mu = B(\bar{\mu}, P)$.

Appendix A reviews the martingale property and other basic properties of the Bayes map.

2 Experimental cost of information

The cost of acquiring information has two main representations: a cost on experiments, as in statistical decision theory, and a cost on random posteriors, as in applications of rational inattention. In this section, we study the relationship between these two representations.

We take the perspective of a decision maker who performs a number of costly experiments, say $n \in \mathbb{N} := \{1, 2, \dots\}$, to gain knowledge about the true value of the state of nature. The experiments are performed *in sequence*: for every $i = 1, \dots, n$, the i -th experiment may depend on the realizations of the first $i - 1$ experiments.

Let X^i be the Cartesian product of i copies of the signal space X , equipped with the product topology. Given a history of signals $x^{i-1} := (x_1, \dots, x_{i-1}) \in X^{i-1}$, the decision maker performs an experiment $P_{x^{i-1}} \in \mathcal{E}$, which stochastically generates a new signal $x_i \in X$. Let $P_i : X^{i-1} \rightarrow \mathcal{E}$ be the map from histories of signals into experiments, which we assume to be Borel measurable; by convention, $X^0 = \{\emptyset\}$. A terminal history is a sequence of n signals $x^n := (x_1, \dots, x_n) \in X^n$. Overall, we represent by $P^n := (P_1, \dots, P_n)$ the decision maker’s experimentation strategy: a *sequential experiment*.

We take a *cost on experiments* $h : \mathcal{E} \rightarrow [0, \infty]$ as primitive; we assume that the function h is Borel measurable. An experiment has an infinite cost if it is not feasible. The decision maker may decide not to experiment at all or to stop experimenting at some point; we include this possibility by assuming that there is an uninformative experiment $P \in \mathcal{E}$ such that $h(P) = 0$.

The decision maker incurs a cost $h(P_{x^{i-1}})$ for performing an experiment $P_{x^{i-1}}$ at a history

x^{i-1} .⁶ Given prior $\pi \in \Delta$, the expected cost of a sequential experiment P^n is

$$h(\pi, P^n) = \sum_{i=1}^n \int_{X^{i-1}} h(P_{x^{i-1}}) dP_\pi^{i-1}(x^{i-1}) \quad (2)$$

where $P_\pi^{i-1} \in \Delta(X^{i-1})$ is the predictive probability generated by π and the first $i - 1$ experiments. By convention, P_π^0 is the Dirac measure concentrated on $x^0 = \emptyset$. Note that the cost of the sequential experiment P^n may depend on the prior of the decision maker—as detailed by (2)—even if the per-period cost is belief independent.

A cost on experiments h induces a cost on random posteriors $c_{h^n} : \Delta^2 \rightarrow [0, \infty]$ by

$$c_{h^n}(\mu) = \inf \{h(\bar{\mu}, P^n) : B(\bar{\mu}, P^n) = \mu\} \quad (3)$$

where $B(\bar{\mu}, P^n)$ is the random posterior generated by a prior $\bar{\mu}$ and a sequential experiment P^n .⁷ The quantity $c_{h^n}(\mu)$ is the least expensive way to generate a random posterior μ by performing a sequential experiment of length n . It is well defined because the set $\{P^n \in \mathcal{E}^n : B(\bar{\mu}, P^n) = \mu\}$ is nonempty, the signal space being rich.⁸

The cost c_{h^n} corresponds to the case in which the decision maker can perform sequential experiments of a fixed length n . We write $c_{h^\infty} : \Delta^2 \rightarrow [0, \infty]$ for the case in which the decision maker can perform sequential experiments of any length:

$$c_{h^\infty}(\mu) = \inf \{c_{h^n}(\mu) : n \in \mathbb{N}\}. \quad (4)$$

Definition 4. A cost on random posteriors $c : \Delta^2 \rightarrow [0, \infty]$ is *experimental* if there is a cost on experiments $h : \mathcal{E} \rightarrow [0, \infty]$ such that $c = c_{h^n}$ for some length $n \in \mathbb{N} \cup \{\infty\}$.

The definition captures the idea of a cost on random posteriors being consistent with a primitive model of costly experimentation. Our first main result is that many cost functions used in rational inattention are *not* experimental. Specifically, many applications of rational inattention consider costs $c_\phi : \Delta^2 \rightarrow [0, \infty)$ that are bounded and *uniformly posterior-separable* (Caplin, Dean and Leahy, 2022):

$$c_\phi(\mu) = \phi(\bar{\mu}) - \int_{\Delta} \phi(p) d\mu(p) \quad (5)$$

where $\phi : \Delta \rightarrow \mathbb{R}$ is a continuous concave function. According to this model, ϕ measures the uncertainty in beliefs, and its expected reduction quantifies the cost of information. A

⁶See Appendix E for a discussion of the case in which the flow cost of information depends on the history of experimentation, for example, because the decision maker builds expertise or experiences fatigue.

⁷To compute $B(\bar{\mu}, P^n)$, treat P^n as a “one-shot” experiment with signal space X^n and follow Section 1.3.

⁸See Lemma 1 in Appendix A.

widespread choice for the uncertainty measure is *entropy* (Matějka and McKay, 2015):

$$\phi_E(p) = - \sum_{\theta} p(\theta) \ln p(\theta). \quad (6)$$

For short, we write c_E instead of c_{ϕ_E} .⁹

Our first main result is that c_{ϕ} is experimental if and only if it is identical to zero:

Proposition 1. *A cost on random posteriors $c_{\phi} : \Delta^2 \rightarrow [0, \infty)$ is bounded, uniformly posterior-separable, and experimental if and only if, for all $\mu \in \Delta^2$, $c_{\phi}(\mu) = 0$.*

Thus, in many cases of interests, posterior-based costs are *inconsistent* with a primitive model of costly experimentation. A predecessor of this result is Mensch (2018, Proposition 4), who, in the language of this paper, shows that $c_{\phi} = c_{h^1}$ for some cost on experiments h if and only if $c_{\phi} = 0$.

The driver behind Proposition 1 is the following property of c_{ϕ} :

Definition 5. A cost $c : \Delta^2 \rightarrow [0, \infty]$ satisfies the *free at full information* (FFI) property if for every state θ ,

$$\lim_{\bar{\mu}(\theta) \rightarrow 1} c(\mu) = 0.$$

In other terms, when the decision maker is almost certain of the state—that is, $\bar{\mu}(\theta) \approx 1$ for some θ —the cost of acquiring information is almost zero—that is, $c(\mu) \approx 0$. It is easy to see that c_{ϕ} satisfies FFI.

FFI seems unappealing in a number of situations; for example, being almost certain that a theorem is true does not necessarily make the proof any easier (think of $P \neq NP$ in computer science). The next proposition shows that FFI is at odds with c being experimental:

Proposition 2. *For every cost on experiment $h : \mathcal{E} \rightarrow [0, \infty]$ and length $n \in \mathbb{N} \cup \{\infty\}$, if the induced cost on random posteriors $c_{h^n} : \Delta^2 \rightarrow [0, \infty]$ satisfies FFI, then there exists a length $m \in \mathbb{N} \cup \{\infty\}$ such that for every prior $\pi \in \Delta$, there exists a sequence of random posteriors (μ_k) in Δ_{π}^2 with*

$$\lim_{k \rightarrow \infty} \mu_k = \sum_{\theta} \pi(\theta) \delta_{\theta}^2 \quad \text{and} \quad \lim_{k \rightarrow \infty} c_{h^m}(\mu_k) = 0$$

In addition, if $n < \infty$, we can choose $m < \infty$; if $n = 1$, we can choose $m = 1$; if Θ is binary, we can choose $m = n$.

A cost on random posteriors c satisfies FFI if, when the prior is dogmatic, information is free. Proposition 2 shows that when we combine FFI with c being experimental, information is free for *every* prior: no matter what their prior π is, a decision maker can perform a

⁹See Denti (2022) for a revealed preference analysis of posterior separable costs.

sufficiently large number m of experiments to gain almost complete knowledge of the state—i.e., $\mu_k \approx \sum_{\theta} \pi(\theta) \delta_{\theta}^2$ —at a negligible cost—i.e., $c_{h^m}(\mu_k) \approx 0$. In sum, FFI and experimental costs coexist only in trivial settings.

The core issue is the relationship between the cost of information and prior beliefs. A broad intuition is that for a generic cost on random posteriors c , the quantity $c(B(\pi, P))$ arbitrarily depends on π . On the other hand, for an experimental cost c_{h^n} , the quantity $c_{h^n}(B(\pi, P))$ depends on π in a structured way, as described by (3) and (4).

We can give a more precise intuition for the case in which $n = 1$ —that is, when the decision maker can perform only one experiment. Suppose that FFI holds and that $c = c_{h^1}$ for some cost on experiments h . Given a prior π and a state θ , FFI implies that for every $\epsilon > 0$, there is $\alpha \in (0, 1)$ sufficiently close to 1 such that

$$c \left(\alpha \delta_{\theta}^2 + (1 - \alpha) \sum_{\tau} \pi(\tau) \delta_{\tau}^2 \right) < \epsilon.$$

The random posterior $\mu_{\alpha} := \alpha \delta_{\theta}^2 + (1 - \alpha) \sum_{\tau} \pi(\tau) \delta_{\tau}^2$ corresponds to complete information for a prior $\pi_{\alpha} := \alpha \delta_{\theta} + (1 - \alpha) \pi$. Since $c = c_{h^1}$, there exists an experiment P_{α} that generates μ_{α} such that

$$c(\mu_{\alpha}) \leq h(P_{\alpha}) < \epsilon.$$

The experiment P_{α} reveals the state *no matter what the prior is* (provided that such a prior is absolutely continuous with respect to π_{α}). Thus, in particular, P_{α} generates $\mu := \sum_{\tau} \pi(\tau) \delta_{\tau}^2$. Since $c = c_{h^1}$,

$$c(\mu) \leq h(P_{\alpha}).$$

We deduce that $c(\mu) < \epsilon$. Since ϵ can be arbitrarily small, we conclude that $c(\mu) = 0$.

The proof for the general case is more involved. The main complication is that the prior affects in a non-trivial way the *expected cost* of a sequential experiment. For example, one would guess that the decision maker could come up with some clever experimentation strategy whose expected cost is small when the prior is dogmatic but substantial when the prior is diffuse. Proposition 2 shows that this intuition, although compelling at first glance, ultimately is misleading.

Section B in the only appendix contains the proofs of Propositions 1 and 2 and the proofs of the other results of the paper; the proof of Proposition 2 comes first since we use the result to derive Proposition 1.

A stark implication of Proposition 2 is that properties of posterior-based costs that may seem technical or innocuous could be quite substantial. For example, let $c : \Delta^2 \rightarrow [0, \infty]$ be a cost on random posteriors that is continuous and such that acquiring no information

is costless—namely, for all $\pi \in \Delta$, $c(\delta_\pi) = 0$. Then, FFI holds:

$$\lim_{\bar{\mu}(\theta) \rightarrow 1} c(\mu) = c(\delta_\theta^2) = 0.$$

Every experimental cost is such that acquiring no information is costless, since the decision maker can always choose not to experiment at all. Thus, a consequence of Proposition 2 is that continuity across random posteriors with different barycenters, a seemingly mild technical assumption, is at odds with c being experimental.

We do not find these results paradoxical. On the contrary, we believe that they show that assumptions on the cost of information across priors require special care. For example, while continuity is at odds with c being experimental, this is not the case for lower semicontinuity:

Proposition 3. *If $h : \mathcal{E} \rightarrow [0, \infty]$ is Blackwell monotone and lower semicontinuous, then $c_{h^1} : \Delta^2 \rightarrow [0, \infty]$ is lower semicontinuous.*

Interestingly, we cannot dispense with the hypothesis that h is Blackwell monotone—see Example 2 in Appendix F.

We emphasize that issues with continuity arise only across priors. In the appendix, we give an example of a cost $c : \Delta^2 \rightarrow [0, \infty)$ that is experimental, non-trivial, lower semicontinuous on Δ^2 , and continuous on each subdomain Δ_π^2 —see Example 3 in Appendix F.

The experiment-based approach puts structure on the relationship between the cost of information and prior beliefs. As discussed above, such structure rules out properties that may seem technical but turn out to be substantial. The flip side is that such structure also delivers properties that may seem strong but are in fact natural. For example, say that a cost on random posteriors $c : \Delta^2 \rightarrow [0, \infty]$ is *concave in the prior* if for all $\alpha \in [0, 1]$, $\pi, \rho \in \Delta$, and $P \in \mathcal{E}$,

$$c(B(\alpha\pi + (1 - \alpha)\rho, P)) \geq \alpha c(B(\pi, P)) + (1 - \alpha)c(B(\rho, P)).$$

Concavity in the prior seems a strong property; at first, it is not obvious why it should hold. The next proposition shows that all experimental costs are concave in the prior.

Proposition 4. *If $c : \Delta^2 \rightarrow [0, \infty]$ is experimental, then c is concave in the prior.*

As observed by Miao and Xing (2020), concavity in the prior has an important implication for information acquisition problems: when a cost on random posteriors is *concave in the prior*, the value function of an information acquisition problem is *convex in the prior*. Miao and Xiang focus on costs that are uniformly posterior-separable; next we extend their

result to arbitrary costs on random posteriors. To this end, we consider the following posterior-based information-acquisition problem:

$$V(\pi) = \max_{\mu \in \Delta_\pi^2} \int_{\Delta} \left(\max_{a \in A} \sum_{\theta} u(a, \theta) p(\theta) \right) d\mu(p) - c(\mu)$$

where π is a prior belief, A is a finite set of actions, and $u : A \times \Theta \rightarrow \mathbb{R}$ is a state-dependent utility.¹⁰ The choice variable is a random posterior $\mu \in \Delta_\pi^2$.

Proposition 5. *If $c : \Delta^2 \rightarrow [0, \infty]$ is concave in the prior, then the value function $V : \Delta \rightarrow \mathbb{R}$ is convex.*

The convexity of the value function has an appealing behavioral interpretation. Consider a two-period information acquisition problem with a persistent state, so that the information acquired “today” can also be used “tomorrow.” Solving the problem backwardly, let $V(\pi)$ be the value of starting tomorrow with prior belief $\pi \in \Delta$ or, equivalently, the value of ending today with posterior belief $p = \pi$. If $V(\pi)$ is a convex function of π , then the prospect of re-using today’s information tomorrow increases the incentive to acquire information today.

A cost function can be concave in the prior but not experimental; for example, the entropy-based cost c_E is concave in the prior (see Cover and Thomas, 2012, Theorem 2.7.4) but not experimental (see Corollary 1). Rational inattention provides an example of a cost that is not concave in the prior. Many applications of rational inattention adopt the specification $c_K : \Delta^2 \rightarrow [0, \infty]$ given by

$$c_K(\mu) = \begin{cases} 0 & \text{if } c_E(\mu) \leq K \\ \infty & \text{otherwise} \end{cases}$$

where $K > 0$ is a bound on the capacity of the decision maker to acquire information. The cost c_K is *quasi-concave* but *not concave* in the prior, as readily checked.

So far, we have presented a number of properties that experimental costs must satisfy. In applications, we may be interested in whether a specific cost function is experimental. The next result helps to answer this question.

Proposition 6. *A cost $c : \Delta^2 \rightarrow [0, \infty]$ satisfies the following properties:*

(i) *If $c = c_{h^1}$ for some $h : \mathcal{E} \rightarrow [0, \infty]$, then $c = c_{h_*^1}$ for $h_* : \mathcal{E} \rightarrow [0, \infty]$ given by*

$$h_*(P) = c(B(\pi^*, P))$$

¹⁰The additive separability of the value of information and the cost of information, which is widespread in the literature, may be justified by underlying quasilinear or CARA preferences.

where π^* is a full-support prior.

(ii) If $c = c_{h^\infty}$ for some $h : \mathcal{E} \rightarrow [0, \infty]$, then $c = c_{h_\star^\infty}$ for $h_\star : \mathcal{E} \rightarrow [0, \infty]$ given by

$$h_\star(P) = \sup_{\pi \in \Delta} c(B(\pi, P)).$$

To check whether a specific cost on random posteriors c is experimental, in principle one would need to go through *all* possible costs on experiments h and check whether $c = c_{h^n}$ for some $n \in \mathbb{N} \cup \{\infty\}$. Proposition 6 simplifies the task in the leading two cases where the decision maker can perform only one experiment (i.e., $n = 1$) or arbitrarily long sequences of experiments (i.e., $n = \infty$). In both cases, it is enough to consider *one* cost on experiments: for $n = 1$, the function h_\star defined by Proposition 6-(i); for $n = \infty$, the function h_\star defined by Proposition 6-(ii).

The simplest case to analyze is $n = 1$. For such a case, in a follow-up paper (Denti, Marinacci and Rustichini, 2022), we provide a complete characterization experimental costs. The case $n = \infty$ is more challenging since it involves a dynamic optimization problem with an infinite horizon and a boundary condition; we hope that our paper can stimulate interest in the problem.

Inter alia, Proposition 6 suggests two simple ways to generate costs on experiments starting from costs on random posteriors. When the starting point is the entropy-based cost, both constructions deliver functional forms that could be of interest for applications, as we explain next.

Given a full-support prior $\pi^* \in \Delta(\Theta)$ —e.g., the uniform prior—we consider the cost on experiments h_{E^*} given by

$$h_{E^*}(P) = c_E(B(\pi^*, P)) \tag{7}$$

where c_E is the entropy-based cost, as defined in (6). Thus, h_{E^*} is the expected reduction in the entropy of beliefs *with respect to a reference prior* π^* .

When the cost is h_{E^*} , optimal information acquisition admits a tractable logit representation that could be of interest for applications. To illustrate, we consider the following experiment-based information-acquisition problem:

$$\max_{P, \beta} \sum_{\theta} \pi(\theta) \int_X \left(\sum_a u(a, \theta) \beta(a|x) \right) dP_{\theta}(x) - h_{E^*}(P) \tag{8}$$

where π is a prior over states, A is a finite set of actions, and $u : A \times \Theta \rightarrow \mathbb{R}$ is a state-dependent utility. The choice variables are an experiment $P : \Theta \rightarrow \Delta(X)$ and a measurable action rule $\beta : X \rightarrow \Delta(A)$, where $\beta(a|x)$ denotes the probability of taking action a after

observing signal x . Let

$$\beta_\theta(a) := \int_X \beta(a|x) dP_\theta(x)$$

be the probability of action a in state θ , and let $\beta_{\pi^*}(a) := \sum_\theta \beta_\theta(a) \pi^*(\theta)$ be the marginal probability of action a with respect to the reference prior π^* .

Proposition 7. *If (P, β) is an optimal solution of (8), then the following conditions hold:*

- For all $\theta \in \Theta$ and $a \in A$,

$$\beta_\theta(a) = \frac{e^{\frac{u(a,\theta)\pi(\theta)}{\pi^*(\theta)}} \beta_{\pi^*}(a)}{\sum_{a'} e^{\frac{u(a',\theta)\pi(\theta)}{\pi^*(\theta)}} \beta_{\pi^*}(a')}.$$

- The probability β_{π^*} is an optimal solution of

$$\max_{\alpha \in \Delta(A)} \sum_\theta \pi^*(\theta) \ln \left(\sum_a e^{\frac{u(a,\theta)\pi(\theta)}{\pi^*(\theta)}} \alpha(a) \right). \quad (9)$$

Conversely, if α is an optimal solution of (9), then there is an optimal solution (P, β) of (8) such that $\beta_{\pi^*} = \alpha$.

The proposition provides an experiment-based version of the logit representation of rationally inattentive behavior by Matějka and McKay (2015), a cornerstone of the literature and a key tool in applications. Matějka and McKay consider an information acquisition problem like (8) but with c_E instead of h_{E^*} for the cost of information. They obtain a logit representation of optimal information acquisition like Proposition 7 but with $\pi = \pi^*$.

Instead of evaluating the cost of an experiment for a fixed reference prior, one could consider the least favorable prior, that is, the prior that makes the experiment most expensive. Specifically, one could consider the cost on experiments h_{E^*} given by

$$h_{E^*}(P) = \max_{\pi \in \Delta} c_E(B(\pi, P)).$$

In information theory, $h_{E^*}(P)$ is called *channel capacity* and is a key measure of information (Cover and Thomas, 2012, Chapter 7). See Woodford (2012) and Nimark and Sundaresan (2019) for applications of channel capacity in economics.

3 Information acquisition in games

The inconsistency of posterior-based costs with a primitive model of costly experimentation becomes concerning when prior beliefs are endogenous, for example, in applications to games.

To illustrate, we revisit the ultimatum game with unobserved offers proposed by Ravid (2020); we discuss more general games in Appendix H.

A seller and a buyer bargain over an indivisible good. The good has value zero for the seller and $v > 0$ for the buyer. The seller makes a take-it-or-leave-it offer $s \in S \subseteq [0, \infty)$ to the buyer. The quantity s is a monetary transfer to the seller if the buyer purchases the good. Ravid assumes that $S = [0, \infty)$. Here, for simplicity, S is finite with $v \in S$ and $\min S < v < \max S$.

In deciding whether or not to purchase the good, the buyer is uncertain about s ; we have in mind settings in which the offer is formulated as a complex contract with many clauses and add-ons. In Ravid (2020) the buyer may also be uncertain about v . Here, for simplicity, v is common knowledge.

To reduce the uncertainty she faces, the buyer can acquire costly information. The buyer’s information acquisition is represented by an experiment $P : S \rightarrow \Delta(X)$. The buyer flexibly chooses how much information to acquire: she can choose any experiment P at a cost. We use the term “experiment” in the broad sense of information structure rather than in the strict sense of statistical procedure. For the buyer, running an experiment could represent reading a long contract, hiring an external consultant, etc. As in Ravid (2020), we assume that the buyer can perform only “one-shot” experiments; the analysis extends to the case in which the buyer can perform multiple experiments in sequence, as we discuss below.

After observing the outcome x of the chosen experiment, the buyer updates her beliefs about s and decides whether or not to purchase the good. A *strategy of the seller* consists of a probability over offers $\sigma \in \Delta(S)$. A *strategy of the buyer* consists of an experiment $P : S \rightarrow \Delta(X)$ and a Borel measurable function $\beta : X \rightarrow [0, 1]$ that specifies for every signal x , the probability $\beta(x) \in [0, 1]$ with which the buyer purchases the good.

With respect to a canonical ultimatum game, the essential friction is that offers are unobserved and information is costly; the cost of an experiment could represent the time and effort to understand a complex contract or the fee paid to a consultant.

Next we analyze the game under two alternative specifications for the cost of information. First, following Ravid (2020), we take a cost on random posteriors as primitive. Then, we revisit his findings by considering a cost on experiments.

Posterior-based costs. We take a cost c on random posteriors as primitive. For running an experiment P , the buyer incurs a cost $c(B(\pi, P))$ where $\pi \in \Delta(S)$ reflects the buyer’s conjecture about the seller’s strategy. In this specification, the cost of information depends on the experiment chosen by the buyer *and* on the buyer’s conjecture about the seller’s strategy; both π and P are endogenous variables.

A strategy σ of the seller is a *best response* to a strategy (P, β) of the buyer if every

$s \in S$ in the support of σ is an optimal solution of the following decision problem:

$$\max_{s'} s' \int_X \beta(x) dP_{s'}(x).$$

A strategy (P, β) of the buyer is a *best response* to a strategy σ of the seller if (P, β) is an optimal solution of the following information acquisition problem:

$$\max_{P', \beta'} \sum_s \sigma(s) \int_X (v - s) \beta'(x) dP'_s(x) - c(B(\sigma, P')).$$

A strategy profile (σ, P, β) is an *equilibrium* if strategies are best responses to each other.

Ravid’s initial finding is a large multiplicity of equilibria: for every $s \in S$ with $s \leq v$, there is an equilibrium (σ, P, β) such that, almost surely, trade happens at price s ; thus, *any* division of the surplus is possible in equilibrium. The result holds under the condition that

$$\text{for all } \pi \in \Delta(S), \quad c(\delta_\pi) = 0. \tag{10}$$

In other terms, c assigns zero cost to every degenerate random posterior that puts probability one on the prior. Virtually all cost functions on random posteriors considered in the literature satisfy (10).

As Ravid (2020, p. 2953) explains, the multiplicity result is driven by the dependence of the cost of information on the buyer’s conjecture about the seller’s strategy. To illustrate, take any $s \in S$ such that $s \leq v$. Suppose that the buyer believes that the seller chooses s with probability one: $\pi(s) = 1$. Since π is degenerate, every experiment induces the same degenerate random posterior that puts probability one on π . By (10), all experiments cost zero. Thus, in particular, the buyer can monitor at zero cost whether or not the seller’s offer is actually s . Then, the buyer can “reward” the seller by purchasing the good if the offer is s and “punish” the seller by not purchasing the good if the offer is different from s .¹¹ For the seller it becomes incentive compatible to offer s with probability one: the buyer’s conjecture about the seller’s strategy is confirmed: $\pi = \sigma$. The buyer ends up purchasing the good at price s without incurring any information cost.

To overcome the multiplicity issue, Ravid proposes a refinement in the spirit of Selten’s trembling-hand perfect equilibrium. Assuming that the cost of information is the expected reduction in the entropy of beliefs, he shows that the refinement is significant: in every equilibrium that survives the refinement, trade is inefficient and, provided that also v is uncertain, the buyer earns a positive payoff (consumer surplus minus attention costs). These results are potentially interesting for they suggest that inattention can be an important

¹¹The punishment can be made sequentially rational by assuming that, off-path, the buyer believes that the seller’s offer is $\max S$.

frictions in negotiations.

As Ravid acknowledges, it is an open question whether or not his refinement has relevance for general posterior-based costs and games.

Experiment-based costs. Ravid (2020) finds that posterior-based costs generate a large multiplicity of equilibria; we agree with him that such equilibria are counterintuitive and not compelling. Ravid argues that the solution concept is problematic and studies a refinement. We believe instead that the problem is with the primitives of the game, in particular, with the specification of the cost of information, as we explain next.

We take a cost on experiments h as primitive; the buyer's information acquisition becomes

$$\max_{P, \beta} \sum_s \sigma(s) \int_X (v - s) \beta(x) dP_s(x) - h(P).$$

The seller's strategy σ affects the value of the information, not the cost of information.

We do not make functional-form assumptions on h ; we just assume that any experiment that carries some information has a positive cost:

$$\text{if } P_s \neq P_{s'} \text{ for some } s, s' \in S, \text{ then } h(P) > 0. \quad (11)$$

We maintain the hypothesis that there exists an uninformative experiment that is free (i.e., the buyer can choose not to engage in any information gathering).

The following proposition shows that the counterintuitive equilibria of Ravid (2020) disappear:

Proposition 8. *Assume (11) for the cost of information. Every equilibrium (σ, P, β) where trade happens with positive probability satisfies the following properties:*

(i) *Trade fails with positive probability:*

$$\sum_s \sigma(s) \int_X \beta(x) dP_s(x) < 1.$$

(ii) *The buyer extracts a positive surplus:*

$$\sum_s \sigma(s) \int_X (v - s) \beta(x) dP_s(x) > 0.$$

(iii) *The seller randomizes between offers below and above v :*

$$\sigma(\{s : s < v\}) > 0 \quad \text{and} \quad \sigma(\{s : s > v\}) > 0.$$

There is a simple intuition behind Proposition 8. Consider, for example, the result that trade fails with positive probability, property (i). Suppose, by contradiction, that trade always occurs. Then the buyer has no incentive to acquire information about the seller's offer, since she purchases the good no matter what. By (11), she must choose not to engage in any information gathering. But this gives an incentive to the seller to deviate and offer $s = \max S$, which makes the good too expensive for the buyer: contradiction.

Under (10), if the buyer believes that the seller's offer is deterministic, then all experiments are free; hence, the buyer can perfectly monitor the seller's offer at no cost, on and off path. This seems odd and is precisely what generates the large multiplicity of equilibria, as Proposition 8 shows. The core issue is the relationship between the cost of information and prior beliefs.

Proposition 8 also shows that an experiment-based approach has practical advantages with the respect to the refinement proposed by Ravid (2020). First, Proposition 8 does not require any functional-form assumption on the cost of information. In addition, the proof of Proposition 8 is short and transparent; Ravid's arguments are clever but also not quite straightforward, since his refinement requires a careful analysis of perturbations of the seller's equilibrium strategy. In Appendix G, we provide an explicit equilibrium characterization for the tractable functional form of the cost of information we describe in (7).

Proposition 8 extends to the case in which $S = [0, \infty)$. It also extends to the case in which the buyer can perform multiple experiments in sequence: provided that the flow cost satisfies (11), Proposition 8 holds since $h(\pi, P^n) = 0$ if and only if P^n is uninformative.

Proposition 8 holds even if the buyer incurs a prior-dependent cost $h(\pi, P)$ for running a one-shot experiment P , as long as (11) holds prior-by-prior: for all $\pi \in \Delta(S)$,

$$\text{if } P_s \neq P_{s'} \text{ for some } s, s' \in S, \text{ then } h(\pi, P) > 0.$$

For example, it could that the cost of P depends on the realized offer or the realized signal. The take-away message is that prior-dependence *per se* is not an issue; the issue we highlight is that such dependence, when not disciplined by a primitive model of costly experimentation, may lead us astray.

Finally, we observe that the experimental cost functions $c = c_{h^n}$ we introduce in this paper also satisfy (10). Thus, if we take c_{h^n} (and not h) as primitive cost, we encounter the same difficulties that arise with the entropy-based cost. Such a modeling choice, however, would not be compelling: c_{h^n} is a derived object: h should be the primitive of the analysis.

4 Conclusion

Over the last two decades, rational inattention has become one of the most influential models of costly information acquisition in economics. Its key ingredient is a rich set of information acquisition strategies, reflecting the fact that in many economic environments, information is abundant but attention is scarce.

Many applications of rational inattention feature a non-trivial dependence of the cost of information on prior beliefs. Such a dependence—somewhat orthogonal to the central idea of flexible information acquisition—becomes concerning in equilibrium analyses where prior beliefs are naturally endogenous. We hope that our work can clarify some of the issues involved and expand the scope of the investigation of information acquisition in economics.

References

- Blackwell, D.**, “Comparison of Experiments,” in “Proc. 2nd Berkeley Symp. on Math. Stat. and Probab.” University of California Press 1951, pp. 93–102.
- and **M. A. Girshick**, *Theory of Games and Statistical Decisions*, John Wiley & Sons, 1954.
- Bloedel, A. W. and W. Zhong**, “The cost of optimally acquired information,” 2021. Available at <https://sites.google.com/site/alexanderbloedel/home>.
- Bohnenblust, H. F., L. S. Shapley, and S. Sherman**, “Reconnaissance in game theory,” 1949. Rand Memorandum 208.
- Caplin, A., M. Dean, and J. Leahy**, “Rationally inattentive behavior: Characterizing and generalizing Shannon entropy,” *Journal of Political Economy*, 2022, *130*, 1676–1715.
- Cover, T. M. and J. A. Thomas**, *Elements of Information Theory*, 2nd ed., John Wiley & Sons, 2012.
- Denti, T.**, “Posterior separable cost of information,” 2022. Available at <https://sites.google.com/site/tommasojdenti/home>.
- , **M. Marinacci, and A. Rustichini**, “The experimental order on random posteriors,” 2022. Available at <https://sites.google.com/site/tommasojdenti/home>.
- Gentzkow, M. and E. Kamenica**, “Costly persuasion,” *American Economic Review P&P*, 2014, *104*, 457–462.
- Gollier, C.**, *The Economics of Risk and Time*, MIT Press, 2001.

- Hébert, B. and M. Woodford**, “Rational inattention when decisions take time,” 2021. Available at <https://bhebert.people.stanford.edu>.
- Le Cam, L.**, “Comparison of experiments: A short review,” *Lecture Notes–Monograph Series*, 1996, *30*, 127–138.
- Maćkowiak, B., F. Matějka, and M. Wiederholt**, “Rational inattention: A review,” 2021. Available at <https://www.aeaweb.org/articles?id=10.1257/jel.20211524&from=f>.
- Matějka, F. and A. McKay**, “Rational inattention to discrete choices: A new foundation for the multinomial logit model,” *American Economic Review*, 2015, *105*, 272–298.
- Mensch, J.**, “Cardinal representations of information,” 2018. Available at <https://sites.google.com/site/jeffreymensch>.
- Miao, J. and H. Xing**, “Dynamic Discrete Choice under Rational Inattention,” 2020. Available at <https://people.bu.edu/miaoj/research.htm>.
- Morris, S. and P. Strack**, “The Wald problem and the equivalence of sequential sampling and static information costs,” 2019. Available at SSRN 2991567.
- Nimark, K. and S. Sundaresan**, “Inattention and belief polarization,” *Journal of Economic Theory*, 2019, *180*, 203–228.
- Pomatto, L., P. Strack, and O. Tamuz**, “The cost of information,” 2020. Available at [arXiv:1812.04211](https://arxiv.org/abs/1812.04211).
- Ravid, D.**, “Ultimatum bargaining with rational inattention,” *American Economic Review*, 2020, *110*, 2948–2963.
- Sims, C. A.**, “Implications of rational inattention,” *Journal of Monetary Economics*, 2003, *50*, 665–690.
- Torgersen, E.**, *Comparison of Statistical Experiments*, Cambridge University Press, 1991.
- Wald, A.**, *Statistical Decision Functions*, John Wiley & Sons, 1950.
- Woodford, M.**, “Inattentive valuation and reference-dependent choice,” 2012. Available at <https://blogs.cuit.columbia.edu/mw2230/>.