INNOCENZO GASPARINI INSTITUTE FOR ECONOMIC RESEARCH

Institutional Members: CEPR, NBER and Università Bocconi

WORKING PAPER SERIES

Experimental Cost of Information

Tommaso Denti, Massimo Marinacci, Aldo Rustichini

Working Paper n. 657

This Version: June 2021

IGIER – Università Bocconi, Via Guglielmo Röntgen 1, 20136 Milano –Italy
http://www.igier.unibocconi.it

Experimental Cost of Information*

Tommaso Denti[1], Massimo Marinacci[2], and Aldo Rustichini[3]

[1]Cornell University
[2]Università Bocconi and IGIER
[3]University of Minnesota

June 19, 2021

Abstract

We relate two 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 the theory of rational inattention. In many cases of interests, the two representations prove to be inconsistent with each other. We provide a systematic analysis of the inconsistency, propose a way around it, and apply our findings to information acquisition in games.

1 Introduction

For Bayesian decision makers, acquisition of information admits two standard representations: a *statistical experiment* $P : \Theta \to \Delta(X)$ 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 relation between these two representations is well understood:[1] 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 $\int p \, \mathrm{d}\mu(p) = \pi$. Moreover, if the set of feasible experiments $\mathcal{E}$ is rich enough, then every random posterior can be induced in this way by some experiment.

Analogously, we can distinguish between two alternative representations for the cost of acquiring information. In one representation, the cost of acquiring information depends

*We are grateful to Jeffrey Mensch and Stephen Morris for useful comments and suggestions. Massimo Marinacci gratefully acknowledges the financial support of ERC (grant INDIMACRO). Aldo Rustichini thanks the National Science Foundation for financial support (grant *SES* 1728056).

[1]See, e.g., the classic Bohnenblust, Shapley, and Sherman (1949) and Blackwell (1951).

on the experiment that the decision maker performs, and is represented by a cost function $h : \mathcal{E} \to [0, \infty]$. This has been the standard representation for the cost of information in statistical decision theory since Wald (1950).

A different perspective has emerged from the work of Sims (2003) on *rational inattention*. In the theory of rational inattention, the cost of acquiring information depends on the random posterior the decision maker ends up with, and is represented by a cost function $c : \Delta(\Delta(\Theta)) \to [0, \infty]$. In a widely adopted specification (e.g., Matějka and McKay, 2015), the quantity $c(\mu)$ is the expected reduction in the entropy of beliefs. Sims's theory of rational inattention has been influential and most recent research on the cost of information has followed in his footsteps. An example is the growing literature on cost functions that are *uniformly posterior separable* (Caplin, Dean, and Leahy, 2021), where entropy is replaced by more general measures of uncertainty.

In this paper we study the relation between the two representations for the cost of information $h : \mathcal{E} \to [0, \infty]$ and $c : \Delta(\Delta(\Theta)) \to [0, \infty]$. Our initial finding is that in many cases of interest, such as the entropy cost, the two representations are inconsistent with each other. The reason is simple: $h(P)$ is a function of the experiment P only, while $c(B(\pi, P))$ potentially varies both with the prior π and the experiment P.

Back to first principles, we define a cost on random posteriors that is consistent with an underlying cost on experiments. Specifically, every cost function h defined on experiments obviously generates a cost function c_h defined on random posterior:

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

where $\bar{\mu} := \int p \, \mathrm{d}\mu(p)$. 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 experiments whose costs are given by h.

Many cost functions c used in applications cannot be rationalized in this way: it may be the case that there is no primitive cost function h defined on experiments that generates c. For instance, the inconsistency arises whenever the cost $c(\mu)$ decreases as the prior $\bar{\mu}$ becomes more dogmatic—that is, as $\bar{\mu}$ puts more probability on a given state being true. This property is satisfied by the entropy cost and, more broadly, by cost functions that are uniformly posterior separable and bounded. These are all examples of cost functions c that are inconsistent with a primitive model of costly experimentation.

This paper introduces the notion of experimental cost of information. A cost function c defined on random posteriors is *experimental* if it is consistent with a primitive model of costly experimentation—that is, if there is a cost function h defined on experiments such that $c = c_h$. Experimental cost functions are our main object of study.

To characterize experimental cost functions, we introduce a new ranking on random posteriors, which we call the *experimental order*. A pair of random posteriors μ and ν is ranked by the experimental order if every experiment that generates μ Blackwell dominates an experiment that generates ν.[2] If $\bar{\mu} = \bar{\nu}$, then μ experimentally dominates ν if and only if μ dominates ν in the convex order, the traditional ranking of random posteriors. The convex order ranks only random posterior with the same mean—that is, with the same underlying prior. The experimental order extends the convex order by allowing comparisons across priors.

We show that a cost c is experimental if and only if it is invariant under $\sim_{ex}$, the symmetric part of the experimental order. If, in addition, c is monotone in the convex order, then c is experimental if and only if it is monotone in the experimental order.

We establish a duality between the class of experimental cost functions c and the class of cost functions h that are invariant under $\sim_b$, the symmetric part of the Blackwell order. Every h that is invariant under $\sim_b$ induces a unique experimental c_h. Conversely, fixing a reference prior π^* (e.g., the uniform prior), every experimental c is induced by a unique h_c that is invariant under $\sim_{ex}$, defined by

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

The duality result is that $c_{h_c} = c$ and $h_{c_h} = h$. If, in addition, c is monotone in the convex order, then h_c is monotone in the Blackwell order; vice versa, if h is Blackwell monotone, then c_h is monotone in the experimental order.

The duality result suggests a regularization scheme to bridge the gap between experimental and non-experimental cost functions. To every cost function c, we can associate an experimental cost function c^* given by

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

We call c^* the *experimental version* of c. We show that the cost function c^* inherits structural properties of the original cost c, such as convexity, lower semicontinuity, and separability.

We apply our findings to information acquisition in games. As shown by Ravid (2020), games with rationally inattentive players present a number of difficulties. In the game he proposes, a seller and a buyer bargain over 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. Ravid adopts the rational inattention model for the buyer's learning problem. He observes that the cost of information depends on the buyer's conjecture about the

[2] The notions used in this introduction are formally defined in Section 3

seller's strategy, an endogenous object, and shows that this generates a large multiplicity of equilibria: any division of the surplus can arise in equilibrium. To overcome the multiplicity issue, Ravid proposes a refinement in the spirit of Selten's perfect equilibrium; the refinement is clever but may sometimes be impractical, for it relies on fine perturbations of equilibrium strategies.

We revisit the bargaining game of Ravid (2020) from the experimental perspective we propose in this paper. We observe that, if the cost of information is experimental, then it is independent of buyer's conjecture about the seller's strategy. We show that, as a result, no equilibrium refinement is needed and the analyst can obtain sharp predictions with transparent arguments and minimal functional-form assumptions on the cost of information. While we focus on a specific game, our methods apply more broadly. Information acquisition in games makes salient the difference between Wald's experimental approach and Sims's rational inattention.

Finally, we extend our framework to sequential information acquisition: the decision maker can perform multiple costly experiments in sequence. We show that the main conclusions of the paper are robust to this extension. When information acquisition is sequential, more cost functions c on random posteriors can be generated, not just functions that are invariant under $\sim_{ex}$. Yet, not every cost function can be generated. In particular, our inconsistency result extends from one-shot to sequential experiments: if the cost $c(\mu)$ decreases as the prior $\bar{\mu}$ becomes more dogmatic, then c is inconsistent with a primitive model of costly experimentation (unless c is identical to zero), regardless of whether experimentation is one-shot or sequential. The findings in the bargaining game are also robust to the possibility that the buyer's learning process is sequential.

1.1 Related literature

Since Sims (2003), most research on the cost of information has focused on functions c defined on random posteriors, rather than on functions h defined on experiments. Among the few exceptions are Mensch (2018) and Pomatto, Strack, and Tamuz (2020).

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 of the Blackwell order 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 uniformly posterior separable and π^* is a fixed reference prior. In the language of our paper, Mensch's representation is the experimental version of c.

To model information production, Pomatto, Strack, and Tamuz (2020) provide an axiomatic foundation to the cost h given by

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

where $\beta(\theta,\tau) \geq 0$, with $\theta,\tau \in \Theta$, and D_{KL} denotes the Kullback-Leibler divergence. To relate their work and rational inattention, they compute the induced cost function on random posteriors, c_h in the language of our paper.

The inconsistency between rational inattention and a primitive model of costly experimentation has been discussed several times in the literature, mostly informally and in specific contexts (see, e.g., Gentzkow and Kamenica, 2014, Mensch, 2018, Nimark and Sundaresan, 2019). Beside 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 $c(B(\pi, P))$ should depend on the prior π or not. A take-home message of our paper is that dependence *per se* is not an issue; for example, if information acquisition is sequential, then the induced cost on random posteriors can exhibit a non-trivial dependence on prior beliefs. The main issue we highlight is that such dependence should be *disciplined*. To put discipline, the simplest way is to assume that information acquisition is one-shot and that c is generated by a primitive cost function h. This benchmark case is the focus of our paper.

In certain settings, it is more reasonable to assume that information acquisition is sequential. Our analysis leaves open the question of what cost functions c on random posteriors can be generated in this way. What we show is that the main conclusions of the paper are robust to sequential learning. The inconsistency between rational inattention and a primitive model of costly experimentation holds regardless of whether experimentation is one-shot or sequential. The findings in the bargaining game extend to the case in which the buyer acquires information sequentially.

A few recent papers have studied the relation between rational inattention and sequential learning (e.g., Hébert and Woodford, 2019; Morris and Strack, 2019; Bloedel and Zhong, 2020). In contrast with our approach, these papers allow the flow cost of information—what would be the function h in our paper—to depend arbitrarily on the evolving beliefs of the decision maker. As a result, they are able to generate a richer class of cost functions on random posteriors; for example, they provide a foundation to the entropy cost and, more broadly, to cost functions that are uniformly posterior separable. Our analysis adds a caveat: it is crucial that the flow cost of information can depend arbitrarily on the decision maker's evolving beliefs; if we put discipline, then the conclusion changes substantially. We postpone a more detailed discussion to Section 7.2.

A second contribution of our work is to show that the difference between Sims's rational inattention and Wald's experimental approach can be of practical relevance. Of course there are settings in which the difference is negligible—e.g., single-agent information acquisition problem where the prior is fixed. The difference becomes salient when prior beliefs are endogenous, such as in games where players acquire information about opponents' past actions. Our analysis highlights the advantages of experimental cost functions to study information acquisition in games. We hope this can encourage more applications; Mensch (2018) already derives interesting predictions on costly monitoring in moral hazard problems.

A third contribution of our work is to provide a bridge between Wald's experimental approach and Sims's rational inattention. The regularization scheme $c \mapsto c^*$ allows to import tools from rational inattention into an experimental framework. For example, as shown by Matějka and McKay (2015), information acquisition problems have a tractable solution for the entropy cost: the behavior of the decision maker admits a simple logit representation. We show that a similar logit representation holds for the experimental version of the entropy cost. The logit representation is one of the main tools in applications of rational inattention; our analysis shows that such tool is available also when the cost of information is experimental.

Overall, we see our work as a positive contribution to the rational inattention literature. Most ingredients of Sims's pioneering work are consistent with an experimental approach: the flexibility in information choice; the role of limited attention; the techniques from information theory.[3] Our goal is to bring these ingredients within the classic framework of statistical decision theory, to facilitate and encourage even more applications of Sims's ideas.

2 Motivating example

Information acquisition in games makes salient the difference between Wald's experimental approach and Sims's rational inattention. To illustrate, we revisit the bargaining game from Ravid (2020) between a seller and a rationally inattentive buyer.

A seller and a buyer bargain over an indivisible good of value $v > 0$. The seller makes a take-it-or-leave-it offer $t \in T \subseteq \mathbb{R}_+$ to the buyer. The quantity t is a monetary transfer to the seller if the buyer purchases the good. Ravid (2020) assumes that $T = \mathbb{R}_+$. Here, for simplicity, we assume instead that T finite with $v \in T$ and $\min T < v < \max T$.

In deciding whether to purchase the good or not, the buyer is uncertain about t; 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 can be uncertain also about v. Here, for

[3]See Maćkowiak, Matějka, and Wiederholt (2021) for a survey of rational inattention, theory and applications.

simplicity, we look at the special case where v is common knowledge.

To reduce the uncertainty she faces, the buyer can acquire information. The buyer's information is represented by an *experiment.* An experiment specifies, for every $t \in T$, a probability distribution P_t over a finite set of signals X. The quantity $P_t(x)$ is the probability that the buyer observes signal x when the seller's offer is t. 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.

The buyer flexibly chooses how much information to acquire: she can choose any experiment $P := (P_t : t \in T)$, that is, any function from T into $\Delta(X)$, the set of probability distributions over X. We assume that X contains at least two distinct signals—otherwise, the information acquisition problem would be trivial. After observing the outcome x of the chosen experiment, the buyer updates her beliefs about t and decides whether to purchase the good or not.

Acquiring 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. We analyze the game under two alternative specifications for the cost of information. First we summarize the results of Ravid (2020) who follows the rational inattention model. Then we revisit his findings from the experimental perspective that we propose in this paper.

2.1 Rational inattention

The cost of information depends on the *random posterior* that the buyer ends up with; it is represented by a function $c : \Delta(\Delta(T)) \to \mathbb{R}_+$. To illustrate, suppose that the buyer believes that the seller randomizes over offers according to $\sigma \in \Delta(T)$. Given an experiment P, the buyer observes a signal x with probability $P_\sigma(x) := \sum_t P_t(x)\sigma(t)$. Provided that $P_\sigma(x) > 0$, Bayesian updating leads to the posterior belief $p_x \in \Delta(T)$ given by $p_x(t) = P_t(x)\sigma(t)/P_\sigma(x)$. The resulting distribution of posterior beliefs, which we denote by $B(\sigma, P)$, assigns probability $P_\sigma(x)$ to p_x. Overall, the buyer incurs cost $c(B(\sigma, P))$ for running experiment P.

In Ravid (2020), the main specification for the cost of information is the expected reduction in the entropy of beliefs:

$$c(B(\sigma, P)) = k \sum_x \left(\sum_t p_x(t) \log p_x(t) \right) P_\sigma(x) - k \sum_t \sigma(t) \log \sigma(t) \tag{1}$$

where $k > 0$ parametrizes the marginal cost of information. The quantity

$$\sum_t p_x(t) \log p_x(t) - \sum_t \sigma(t) \log \sigma(t) = \left(-\sum_t \sigma(t) \log \sigma(t) \right) - \left(-\sum_t p_x(t) \log p_x(t) \right)$$

is the difference between the entropies of σ and p_x. Being entropy a measure of uncertainty, entropy reduction provides an estimate of the amount of information gathered by the buyer. Entropy reduction is a widely adopted specification for the cost of information in rational inattention (see, e.g., Matějka and McKay, 2015).[4]

A *strategy of the seller* consists of a probability distribution over offers $\sigma \in \Delta(T)$. A *strategy of the buyer* consists of an experiment P and a function $\beta : X \to [0, 1]$ that specifies, for every signal x, the probability $\beta(x) \in [0, 1]$ with which the buyer purchases the good. A strategy σ of the seller is a *best response* to a strategy (P, β) of the buyer if, for every $t \in T$ such that $\sigma(t) > 0$,

$$\sum_x t\beta(x)P_t(x) = \max_{t'} \sum_x t'\beta(x)P_{t'}(x).$$

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 information acquisition problem

$$\max_{P', \beta'} \sum_{t,x} (v - t)\beta'(x)P'_t(x)\sigma(t) - c(B(\sigma, P')).$$

A strategy profile (σ, P, β) is an *equilibrium* if strategies are best responses to each other.[5]

The initial finding of Ravid (2020) is a large multiplicity of equilibria:

Proposition 1 (Ravid, 2020)**.** *Assume (1) for cost of information. For every $t \in T$ with $t \le v$, there is an equilibrium (σ, P, β) such that, almost surely, trade happens at price t:*

$$\sigma(t) = 1 \qquad \text{and} \qquad \sum_x \beta(x)P_t(x) = 1.$$

Thus, *any* division of the surplus can arise in equilibrium. As Ravid (2020, p. 2953) explains, the result is driven by the dependence of the cost of information $c(B(\sigma, P))$ on the buyer's conjecture about the seller's strategy σ. To illustrate, take any $t \in T$ with $t \le v$. Suppose that the buyer believes that seller chooses t with probability one: $\sigma(t) = 1$. Since

[4] In (1), the quantity $c(B(\sigma, P))$ is a concave function of σ. As a result, the buyer's cost of information depends on her beliefs about the seller's strategy in a non-linear fashion. As explained by Ravid (2020, p. 2952), this non-linearity makes the model *a psychological game* (Geanakoplos, Pearce, and Stacchetti, 1989). In the experimental approach we propose in this paper, the buyer's cost of information depends only on P; thus the model falls within the traditional theory of games.

[5] In the definition of equilibrium, Ravid (2020) includes the buyer's beliefs about the seller's strategy and adds the condition that, in equilibrium, they must be consistent with the seller's actual strategy. Here, for simplicity, we omit the buyer's beliefs from the definition of equilibrium.

σ is degenerate, every experiment induces the same degenerate random posterior that puts probability one on σ. By (1), all experiments cost zero. Thus, in particular, the buyer can monitor at zero cost whether the seller's offer is actually t or not. The buyer then can "reward" the seller by purchasing the good if the offer is t and "punish" the seller by not purchasing the good if the offer is different from t.[6] For the seller it becomes incentive compatible to offer t: the buyer's conjecture about the seller's strategy is confirmed. The buyer ends up purchasing the good at price t without paying any information cost.

To get around the multiplicity issue, Ravid (2020) proposes an equilibrium refinement in the spirit of Selten's trembling-hand perfect equilibrium. Ravid deems an equilibrium (σ, P, β) *credible* if σ is the limit in $\Delta(T)$ of a sequence (σ_n) such that, for every n, σ_n has full support and (P, β) is a best response to σ_n.[7] Ravid shows that the refinement has substantial bite:

Proposition 2 (Ravid, 2020)**.** *Assume (1) for cost of information. If (σ, P, β) is a credible equilibrium, then trade fails with positive probability:*

$$\sum_x \beta(x) P_t(x) \sigma(t) < 1.$$

If $k < v$, then there exists an equilibrium where trade happens with positive probability; in such equilibrium,

$$\sigma(v) = 1 \qquad \textit{and} \qquad \sum_x \beta(x) P_v(x) = 1 - \frac{k}{v}.$$

Thus, in every credible equilibrium, trade is inefficient. To appreciate the result, consider the benchmark case without inattention where the buyer perfectly observes the seller's offer. Without inattention, the game has an unique (subgame perfect) equilibrium in which the sellers offers v and the buyer accepts any offer $t \leq v$, rejects otherwise; in particular, trade is efficient.[8] Ravid's result shows that inattention reverses this conclusion: if the buyer is inattentive, trade is inefficient in every credible equilibrium.

The inefficiency of trade in credible equilibria is the first main result of Ravid (2020). The result extends to the case in which the buyer is uncertain about v. The second main result of Ravid (2020) applies specifically to the case in which the buyer is uncertain about v. If the buyer is inattentive both about v and about t, Ravid shows that, in every credible equilibrium where trade happens with positive probability, the buyer extracts a positive

[6] The punishment can be made sequentially rational by assuming that off-path the buyer believes that the seller's offer is $\max T$.

[7] In Ravid (2020), the definition of credible equilibrium is slightly more involved because T is infinite and this introduces some measure-theoretic subtleties.

[8] This is the unique subgame perfect equilibrium when $T = \mathbb{R}_+$. When T is discrete, there could be other subgame perfect equilibria.

surplus from the seller. As detailed by Proposition 2, the result does not hold when v is common knowledge. When v is common knowledge, the seller offers $t = v$ with probability one, which implies that the buyer's surplus is zero.[9]

Ravid shows that his results extend to the case in which entropy in (1) is replaced by a more general measure of uncertainty:

$$c(B(\sigma, P)) = \sum_x \phi(p_x)P_\sigma(x) - \phi(\sigma) \tag{2}$$

where $\phi : \Delta(T) \to \mathbb{R}$ is continuous and strictly convex. As Ravid (2020, p. 2962) acknowledges, his analysis leaves open the question whether the results extend to more general cost functions or not.

2.2 Experimental approach

The cost of information $h(P) \in \mathbb{R}_+$ depends only on the experiment P that the buyer chooses. We do not make functional form assumptions on h; we just assume that an experiment costs zero if and only if is uninformative:

$$h(P) = 0 \text{ if and only if } P_t = P_{t'} \text{ for all } t, t' \in T. \tag{3}$$

We define equilibrium to parallel Ravid (2020). A strategy profile (σ, P, β) is an *equilibrium* if the following conditions hold:

- for every $t \in T$ such that $\sigma(t) > 0$,

$$\sum_x t\beta(x)P_t(x) = \max_{t'} \sum_x t'\beta(x)P_{t'}(x).$$

- (P, β) is an optimal solution of

$$\max_{P',\beta'} \sum_{t,x} (v - t)\beta'(x)P'_t(x)\sigma(t) - h(P').$$

The next result revisits the findings of Ravid (2020) from the experimental perspective we propose in this paper.

Proposition 3. *Assume (3) for cost of information. If (σ, P, β) is an equilibrium, then*

[9] This is the unique credible equilibrium when $T = \mathbb{R}_+$. When T is discrete, there could be other credible equilibria.

trade fails with positive probability:

$$\sum_x \beta(x) P_t(x) \sigma(t) < 1.$$

Moreover, if trade happens with positive probability, then the buyer extracts a positive surplus:

$$\sum_x \beta(x) P_t(x) \sigma(t) > 0 \quad \Longrightarrow \quad \sum_{t,x} (v-t) \beta(x) P_t(x) \sigma(t) > 0;$$

In particular, the seller randomizes between offers below and above v:

$$\sum_x \beta(x) P_t(x) \sigma(t) > 0 \quad \Longrightarrow \quad \max_{t<v} \sigma(t) > 0 \quad \textit{and} \quad \max_{t>v} \sigma(t) > 0.$$

Proof. First, assume by contradiction that (σ, P, β) is an equilibrium where trade happens with probability one:

$$\sum_x \beta(x) P_t(x) \sigma(t) = 1.$$

By (3), the experiment P must be uninformative; otherwise, the buyer would have a profitable deviation (P', β') where P' is uninformative and, for every signal x, $\beta'(x) = 1$. Since P is uninformative, $\sum_x \beta(x) P_t(x) = 1$ for all $t \in T$. Thus σ must put probability one on $t = \max T$; otherwise, the seller would have a profitable deviation in increasing his offer. Since $\max T > v$, the buyer must never purchase the good: $\beta(x) = 0$ for all x. This contradicts the hypothesis that trade happens with probability one. We conclude that in every equilibrium trade fails with positive probability.

Next, assume by contradiction that (σ, P, β) is an equilibrium where trade happens with positive probability and the buyer extracts zero surplus:

$$\sum_x \beta(x) P_t(x) \sigma(t) > 0 \qquad \text{and} \qquad \sum_{t,x} (v-t) \beta(x) P_t(x) \sigma(t) = 0.$$

By (3), P must be uninformative; otherwise, the buyer would have a profitable deviation (P', β') where P' is uninformative and, for every signal x, $\beta'(x) = 0$. Since P is uninformative, $\sum_x \beta(x) P_t(x) = \sum_x \beta(x) P_{t'}(x) > 0$ for all $t, t' \in T$. Thus σ must put probability one on $t = \max T$: otherwise, the seller would have a profitable deviation in increasing his offer. Since $\max T > v$, the buyer must never purchase the good: $\beta(x) = 0$ for all x. This contradicts the hypothesis that trade happens with positive probability. We conclude that, in every equilibrium where trade happens with positive probability, the buyer extracts a positive surplus. It follows immediately that $\sigma(t) > 0$ for some $t < v$.

Finally, assume by contradiction that (σ, P, β) is an equilibrium where trade happens with positive probability and the seller's offer is never above v:

$$\sum_x \beta(x) P_t(x) \sigma(t) > 0 \qquad \text{and} \qquad \max_{t > v} \sigma(t) = 0.$$

By (3), the experiment P must be uninformative; otherwise, the buyer would have a profitable deviation (P', β') where P' is uninformative and, for every signal x, $\beta'(x) = 1$. Since P is uninformative, $\sum_x \beta(x) P_t(x) = \sum_x \beta(x) P_{t'}(x) > 0$ for all $t, t' \in T$. Thus σ must put probability one on $t = \max T$: otherwise, the seller would have a profitable deviation in increasing his offer. This contradicts the hypothesis that the seller's offer is never above v. We conclude that, in every equilibrium where trade happens with positive probability, $\sigma(t) > 0$ for some $t > v$. □

The proposition validates the findings of Ravid (2020), but also shows the advantages of the experimental approach. First, Proposition 3 holds for all equilibria, not just for credible equilibria. Second, Proposition 3 does not require any functional form assumption on the cost of information. Third, the proof of Proposition 3 is short and transparent; the arguments proposed by Ravid (2020) are clever but also not quite straightforward, since they require a careful analysis of perturbations of the seller's equilibrium strategy. Fourth, the first two statements of Proposition 3 and their proofs extend verbatim to the case where the buyer does not know v; thus we do not see the somewhat artificial distinction between the cases of known and unknown v that arises in Ravid (2020). When v is common knowledge, in every credible equilibrium of Ravid (2020), the seller offers $t = v$ with probability one. By contrast, per the last statement of Proposition 3, in every equilibrium of our revisitation the seller randomizes between offers below and above v.[10]

This example and, more broadly, information acquisition in games motivate the rest of the paper, where we provide a systematic analysis of the relation between Wald's experimental approach and Sims's rational inattention.

3 Setup

We review a few preliminary notions, most of them due to Bohnenblust, Shapley, and Sherman (1949) and Blackwell (1951).[11]

[10] Proposition 3 extends verbatim to the case where $T = [0, \infty)$ as in Ravid (2020).

[11] An early monograph covering statistical experiments is Blackwell and Girshick (1954), a more recent one is Torgersen (1991). Le Cam (1996) authoritatively reviews the topic.

3.1 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 denote by Δ_+ the set of probabilities on Θ with full support. We will often use the uniform prior as reference point, which we denote by π^*. In a few examples, we work with a binary state space $\Theta = \{0, 1\}$. When the state is binary, we identify Δ and the unit interval $[0, 1]$ under the convention that $\pi \in [0, 1]$ is the probability that $\theta = 1$.

Let Δ^2 be the set of Borel probabilities on Δ. Its elements, generically denoted by μ and ν, will be interpreted as *random posteriors*, that is, as probability distributions over posterior beliefs. We use the symbol δ_π for the Dirac measure concentrated on π. We endow Δ^2 with the weak* topology: a sequence (μ_n) converges to μ if $\int \phi \, \mathrm{d}\mu_n \to \int \phi \, \mathrm{d}\mu$ for every continuous function $\phi : \Delta \to \mathbb{R}$.

The probability $\bar{\mu}$ over states defined by

$$\bar{\mu} = \int_\Delta p \, \mathrm{d}\mu(p)$$

is the *barycenter* of μ. It will be interpreted as the prior from which μ is obtained via Bayesian updating. Let Δ^2_π be the set of random posteriors with barycenter π. Let Δ^2_+ be the union of all Δ^2_π such that π has full support.

Elements of Δ^2 are ranked by the convex order. Let $Cv(\Delta)$ be the set of functions $\phi : \Delta \to \mathbb{R}$ 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 $\phi \in Cv(\Delta)$, the inequality $\int \phi \, \mathrm{d}\mu \geq \int \phi \, \mathrm{d}\nu$ holds.

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

3.2 Experiments

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

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.

An experiment P is *simple* if all P_θ have finite support. For simple experiments we use the symbol $\operatorname{supp} P$ for the union of the supports of the P_θ:

$$\operatorname{supp} P = \{x \in X : P_\theta(x) > 0 \text{ for some } \theta\}.$$

We call $\operatorname{supp} P$ the *support* of P.

We endow $\mathcal{E}$ with a statewise mixture operation defined, for all $\alpha \in [0,1]$, by

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

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 product topology: a sequence of experiments (P_n) converges to an experiment P if, for every state θ, the sequence $(P_{n,\theta})$ converges to P_θ in $\Delta(X)$.

Experiments whose signal space is Δ are called *standard*. Unless stated otherwise, we assume that X is rich enough to embed Δ. Formally, we assume that the set 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. This richness assumption allows us to identify the set of standard experiments with a subset of $\mathcal{E}$.

It will be useful to think of experiments in terms of likelihood ratios. Being Θ finite, for every experiment P we can find a control (σ-finite) measure λ such that all P_θ are absolutely continuous with respect to λ.. The corresponding family of densities is

$$\frac{\mathrm{d}P}{\mathrm{d}\lambda} = \left(\frac{\mathrm{d}P_\theta}{\mathrm{d}\lambda}\right)_{\theta\in\Theta}.$$

We adopt the conventions $0/0 = 0$ and $0\cdot\infty = 0$. When no confusion should arise, we avoid the "almost surely" quantifier.

Example 1. For a simple experiment P, the control measure λ can simply be the counting measure on the support: $\lambda(x) = 1$ if $x \in \operatorname{supp} P$ and $\lambda(x) = 0$ otherwise. If so, then $\mathrm{d}P_\theta(x)/\mathrm{d}\lambda(x) = P_\theta(x)$ for all $x \in \operatorname{supp} P$. If P is not simple, we can pick $\lambda = \sum_\theta P_\theta$. In terms of likelihood ratios, the particular choice of λ is inconsequential. ▲

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

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

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 $K : X \to \Delta(X)$ such that $Q_\theta = KP_\theta$ for all θ.

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

3.3 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*

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

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 likelihood of different signal realizations. The process of Bayesian updating associates, for P_π-almost all x, a posterior belief $p_x \in \Delta$ given by

$$p_x(\theta) = \frac{\mathrm{d}P_\theta(x)/\mathrm{d}\lambda(x)}{\sum_\tau \pi(\tau)\,(\mathrm{d}P_\pi(x)/\mathrm{d}\lambda(x))}.$$

When P is simple, then p_x follows the usual Bayes rule: for all x such that $P_\pi(x) > 0$,

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

The densities $\mathrm{d}P_\theta/\mathrm{d}\lambda$ allow to extend Bayes rule beyond simple experiments that can generate at most finitely many signals.

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

$$B(\pi, P) = \sum_{x \in \operatorname{supp} P_\pi} P_\pi(x)\, \delta_{p_x}$$

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 $\mu = B(\bar{\mu}, P)$.

We denote by P^μ a generic experiment that induces μ. Section A in the appendix reviews the martingale property and other basic properties of the Bayes map.

4 Analysis

We introduce experimental cost functions, our main object of study. We take the perspective of a decision maker who can perform an experiment $P \in \mathcal{E}$ at a cost $h(P) \in [0, \infty]$. An infinite costs represents an experiment that is not feasible. The decision maker may decide not to perform any experiment; we represent the possibility by assuming that $h(P) = 0$ for some uninformative experiment P.

The primitive cost $h : \mathcal{E} \to [0, \infty]$ induces a cost on random posteriors $c_h : \Delta^2_+ \to [0, \infty]$ defined by

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

The quantity $c_h(\mu)$ represents the least expensive way to achieve the random posterior μ for a decision maker with prior $\bar{\mu}$ by performing experiments whose costs are represented by the function h. It is well-defined because the set $\{P : B(\bar{\mu}, P) = \mu\}$ is nonempty (being $\mathcal{E}$ rich enough; see Lemma 7 in the appendix).

Definition 4. A cost function $c : \Delta^2_+ \to [0, \infty]$ is *experimental* if it is induced by some primitive cost function $h : \mathcal{E} \to [0, \infty]$, i.e., $c = c_h$.

To simplify the exposition, we restrict the domain of c to random posteriors in Δ^2_+. The extension to priors with partial support presents no difficulties, as we discuss in Section C in the online appendix.

In the rest of the section, we characterize what cost functions on random posteriors are experimental. Our characterization is based on a new order on random posteriors:

Definition 5. The *experimental order* $\succeq_{ex}$ is a binary relation on Δ^2 defined by $\mu \succeq_{ex} \nu$ if, for every $P \in \mathcal{E}$ such that $B(\bar{\mu}, P) = \mu$, there is $Q \in \mathcal{E}$ such that $P \succeq_b Q$ and $B(\bar{\nu}, Q) = \nu$.

In words, μ experimentally dominates ν when every experiment that generates the random posterior μ Blackwell dominates an experiment that generates the random posterior ν. Observe that the barycenters $\bar{\mu}$ and $\bar{\nu}$ may be different. Indeed, as the next lemma shows, the experimental order is a weakening of the convex order that permits to compare random posteriors with different barycenters.

Lemma 1. *The experimental order is a preorder. Moreover,*

(i) $\mu \succeq_{ex} \nu$ *if and only if*

$$\{P \in \mathcal{E} : B(\bar{\mu}, P) \succeq_{cv} \mu\} \subseteq \{Q \in \mathcal{E} : B(\bar{\nu}, Q) \succeq_{cv} \nu\};$$

(ii) $\mu \sim_{ex} \nu$ *if and only if*

$$\{P \in \mathcal{E} : B(\bar{\mu}, P) = \mu\} = \{Q \in \mathcal{E} : B(\bar{\nu}, Q) = \nu\};$$

(iii) if $\mu \succeq_{cv} \nu$*, then* $\mu \succeq_{ex} \nu$*. The converse holds if* $\bar{\mu} = \bar{\nu}$*.*

The relation between $\succeq_{ex}$ and $\succeq_{cv}$ is characterized by (i), which also provides an alternative definition for the experimental order: $\mu \succeq_{ex} \nu$ when every experiment that induces a random posterior more dispersed than μ, also induces a random posterior more dispersed than ν, that is,

$$B(\bar{\mu}, P) \succeq_{cv} \mu \Longrightarrow B(\bar{\nu}, P) \succeq_{cv} \nu \qquad \forall P \in \mathcal{E}.$$

The special case where $\mu \sim_{ex} \nu$ is characterize by (ii): $\mu \sim_{ex} \nu$ if and only if μ and ν are generated by the same experiments. Overall, the experimental order is a weakening of the convex order, but the two rankings coincide on random posteriors with the same barycenter, as detailed in (iii). The next example provides a simple description of these relations.

Example 2. For every $\theta \in \Theta$, let $\delta_{\delta_\theta} \in \Delta^2$ be the Dirac measure concentred on the Dirac measure $\delta_\theta \in \Delta$ that puts probability one on the state being θ. For all $\pi \in \Delta_+$ and $\mu \in \Delta^2$, $\sum_\theta \pi(\theta)\delta_{\delta_\theta} \succeq_{ex} \mu$. If $\bar{\mu} = \pi$, then $\sum_\theta \pi(\theta)\delta_{\delta_\theta} \succeq_{cv} \mu$. If $\bar{\mu} \neq \pi$, then $\sum_\theta \pi(\theta)\delta_{\delta_\theta} \not\succeq_{cv} \mu$. If $\sum_\theta \pi(\theta)\delta_{\delta_\theta} \sim_{ex} \mu$, then $\mu = \sum_\theta \rho(\theta)\delta_{\delta_\theta}$ for some $\rho \in \Delta_+$. ▲

Next we relate experimental and Blackwell orders.

Lemma 2. *For any* $P, Q \in \mathcal{E}$*, the following conditions are equivalent:*

(i) $P \succeq_b Q$*;*

(ii) $B(\pi, P) \succeq_{ex} B(\rho, Q)$ *for all* $\pi, \rho \in \Delta_+$*;*

(iii) $B(\pi, P) \succeq_{ex} B(\rho, Q)$ *for some* $\pi, \rho \in \Delta_+$*.*

The experimental order has thus a simple Blackwell characterization when priors have full support. This implies, *inter alia*, a variational representation of the experimental order in terms of sublinear functions. Here $Cs(\mathbb{R}^\Theta_+)$ denotes the collection of continuous sublinear functions $\psi : \mathbb{R}^\Theta_+ \to \mathbb{R}$. To shorten notation, we denote by $p/\bar{\mu}$ the vector $(p(\theta)/\bar{\mu}(\theta))_{\theta \in \Theta}$.

Lemma 3. *For all* $\mu, \nu \in \Delta^2_+$*, we have* $\mu \succeq_{ex} \nu$ *if and only if*

$$\int_\Delta \psi\left(\frac{p}{\bar{\mu}}\right) \mathrm{d}\mu(p) \geq \int_\Delta \psi\left(\frac{p}{\bar{\nu}}\right) \mathrm{d}\nu(p) \qquad \forall \psi \in Cs(\mathbb{R}^\Theta_+).$$

The result extends to the case in which $\psi : \Delta \to (-\infty, \infty]$ is sublinear and lower semicontinuous.

The next theorem relates experimental cost functions to the experimental order. Let $\mathcal{H}$ be the class of cost functions $h : \mathcal{E} \to [0, \infty]$ such that $h(P) = 0$ for some $P \in \mathcal{E}$; let $\mathcal{C}$ be the class of cost functions $c : \Delta_+^2 \to [0, \infty]$ such that $c(\delta_\pi) = 0$ for all $\pi \in \Delta_+$.

Theorem 1. *(i) A cost function $c \in \mathcal{C}$ is experimental if and only if it is invariant under $\sim_{ex}$, that is,*

$$\mu \sim_{ex} \nu \Longrightarrow c(\mu) = c(\nu) \qquad \forall \mu, \nu \in \Delta_+^2.$$

Moreover, c is induced by a unique $h_c \in \mathcal{H}$ that is invariant under $\sim_b$, given by

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

(ii) If $h \in \mathcal{H}$ is monotone in the Blackwell order, then c_h is monotone in the convex order and

$$c_h(\mu) = \inf\{h(P) : B(\bar{\mu}, P) \succeq_{cv} \mu\} \qquad \forall \mu \in \Delta_+^2.$$

(iii) A cost function $c \in \mathcal{C}$ is monotone in the convex order and experimental if and only if it is monotone in the experimental order. Moreover, h_c is Blackwell monotone.

As the proof in the appendix shows, the choice of the uniform prior π^* is a convenient normalization. We could have chosen any other prior with full support.

In many specifications of rational inattention, the cost function $c : \Delta_+^2 \to [0, \infty]$ is not invariant under $\sim_{ex}$. Thus, by Theorem 1, it is not experimental. For example, it is often assumed that $c(\mu)$ decreases as the underlying prior $\bar{\mu}$ becomes more dogmatic. A corollary of Theorem 1 is that such assumption is satisfied by only one experimental cost function, the trivial cost function.

Corollary 1. *For every experimental cost $c : \Delta_+^2 \to [0, \infty]$, the following conditions are equivalent:*

(i) *For some $\theta \in \Theta$, $\lim_{\bar{\mu}(\theta) \to 1} c(\mu) = 0$.*

(ii) *For all $\mu \in \Delta_+^2$, $c(\mu) = 0$.*

Most notably, (i) is satisfied by the widespread entropy cost

$$c_R(\mu) := \int_\Delta \left(\sum_\theta p(\theta) \log p(\theta) \right) \mathrm{d}\mu(p) - \sum_\theta \bar{\mu}(\theta) \log \bar{\mu}(\theta). \tag{4}$$

See also (1). By Corollary 1, the cost function c_R is not experimental, being c_R not identical to zero. More broadly, (i) is satisfied by cost functions that are *uniformly posterior separable*

(Caplin, Dean, and Leahy, 2021) and bounded: given $\phi \in Cv(\Delta)$,[12]

$$c_\phi(\mu) = \int_\Delta \phi(p)\,\mathrm{d}\mu(p) - \phi(\bar{\mu}). \tag{5}$$

See also (2). By Corollary 1, a cost function c is experimental, uniformly posterior separable, and bounded if and only if it is identical to zero. A related result is Mensch (2018, Proposition 4), which shows that the quantity $\int \phi \,\mathrm{d}B(\pi, P) - \phi(\pi)$ is constant in π if and only if ϕ is affine (and therefore $\int \phi \,\mathrm{d}B(\pi, P) = \phi(\pi)$).

Theorem 1 suggests a duality between functions that are invariant under $\sim_b$ and $\sim_{ex}$. Let $\mathcal{H}^b \subseteq \mathcal{H}$ be the class of functions $h : \mathcal{E} \to [0, \infty]$ that are invariant under $\sim_b$, and let $\mathcal{C}^{ex} \subseteq \mathcal{C}$ be the class of functions $c : \Delta^2_+ \to [0, 1]$ that are invariant under $\sim_{ex}$.

Corollary 2. *The map $D : \mathcal{H}^b \to \mathcal{C}^{ex}$ defined by $D(h) = c_h$ is bijective. Its inverse map $D^{-1} : \mathcal{C}^{ex} \to \mathcal{H}^b$ is given by $D^{-1}(c) = h_c$. Moreover, if h is Blackwell monotone, then $D(h)$ is monotone in the experimental order; if c is monotone in the experimental order, than $D^{-1}(c)$ is Blackwell monotone.*

We can diagram the duality, which is at the heart of our paper, as follows:

$$\begin{array}{cccc} & h \xrightarrow{\quad D \quad} c_h & & \\ \mathcal{H}^b & & & \mathcal{C}^{ex} \\ & h_c \xleftarrow{\quad D^{-1} \quad} c & & \end{array}$$

The duality suggests a regularization scheme to bridge the gap between experimental and non-experimental cost functions. To illustrate, observe first that the nature of an experimental cost function is determined by its restriction on $\Delta^2_{\pi^*}$, the set of random posteriors with barycenter π^*. For a pair of experimental cost functions c and c', if $c(\mu) = c'(\mu)$ for all $\mu \in \Delta^2_{\pi^*}$, then by definition $h_c = h_{c'}$, which in turn implies $c = c'$ by the duality result (Corollary 2). This observation suggests a normalization map for random posteriors.

Lemma 4. *There is a map $\Delta^2_+ \ni \mu \mapsto \mu^* \in \Delta^2_{\pi^*}$ such that $\mu \sim_{ex} \mu^*$. In particular, the following properties are satisfied:*

(i) *$\nu \succeq_{ex} \mu$ if and only if $\nu^* \succeq_{cv} \mu^*$;*

(ii) *$c(\mu) = c(\mu^*)$ whenever c is experimental.*

The normalization map $\mu \mapsto \mu^*$ assigns to every random posterior $\mu \in \Delta^2_+$ the unique element of the equivalence class $[\mu]_{ex}$ whose barycenter is π^*. The next example provides an explicit expression for μ^*.

[12]In the unbounded case, ϕ may take infinite values outside Δ_+.

Example 3. Let $\mu \in \Delta^2_+$ have finite support $\{p_1, \dots, p_n\}$. The random posterior μ is generated by an experiment P, with finite support $\{x_1, \dots, x_n\}$, given by

$$P_\theta(x_i) = \frac{p_i(\theta)\mu(p_i)}{\bar{\mu}(\theta)}.$$

The normalization μ^* satisfies $\mu^* = B(\pi^*, P)$. Thus μ^* has finite support $\{p^*_1, \dots, p^*_n\}$ with

$$p^*_i(\theta) = \frac{P_\theta(x_i)\pi^*(\theta)}{P_{\pi^*}(x_i)} = \frac{p_i(\theta)(\pi^*(\theta)/\bar{\mu}(\theta))}{\sum_\tau p_i(\tau)(\pi^*(\tau)/\bar{\mu}(\tau))}.$$

In addition,

$$\mu^*(p^*_i) = P_{\pi^*}(x_i) = \sum_\theta \frac{p_i(\theta)\pi^*(\theta)}{\bar{\mu}(\theta)}\mu(p_i).$$

The expressions generalize to the case where μ does not have finite support, using Radon-Nikodym derivatives.[13] ▲

The map $\mu \mapsto \mu^*$ permits to introduce a regularization scheme for cost functions that are not experimental.

Corollary 3. *Given $c \in \mathcal{C}$, the function $c^* : \Delta^2_+ \to [0, \infty]$ defined by*

$$c^*(\mu) = c(\mu^*)$$

is experimental. If, in addition, c is monotone in convex order, then c^* is monotone in the experimental order.

We call c^* the *experimental version* of c. We can enrich the last diagram as follows:

$$\begin{array}{ccccc} & h \xrightarrow{\;D\;} c_h & & c^* \longleftarrow c & \\ \mathcal{H}^b & & \mathcal{C}^{ex} & & \mathcal{C} \\ & h_c \xleftarrow{\;D^{-1}\;} c & & & \end{array}$$

In the next examples, we regularize the costs c_R and c_ϕ defined in (4) and (5).

Example 4. Let $D_{KL}(\xi_1 \| \xi_2)$ be the Kullback-Leibler divergence of $\xi_1, \xi_2 \in \Delta(X)$:

$$D_{KL}(\xi_1\|\xi_2) = \begin{cases} \int_X \log\left(\mathrm{d}\xi_1/\mathrm{d}\xi_2\right)\, \mathrm{d}\xi_1 & \text{if } \xi_1 \ll \xi_2, \\ \infty & \text{otherwise.} \end{cases}$$

[13] Define $\nu \in \Delta^2_+$ by $\mathrm{d}\nu(p) = \sum_\theta \frac{p(\theta)\pi^*(\theta)}{\bar{\mu}(\theta)}\mathrm{d}\mu(p)$. Then μ^* is the pushforward of ν under the map

$$p \longrightarrow \frac{p(\theta)(\pi^*(\theta)/\bar{\mu}(\theta))}{\sum_\tau p(\tau)(\pi^*(\tau)/\bar{\mu}(\tau))}.$$

The cost function c_R can be written as an average of Kullback-Leibler divergences:

$$c_R(\mu) = \sum_\theta D_{KL}(P_\theta^\mu \| P_\pi^\mu)\pi(\theta)$$

where P^μ is an experiment that generates μ. The experimental version of c_R is

$$c_R^*(\mu) = c_R(\mu^*) = \sum_\theta D_{KL}(P_\theta^\mu \| P_{\pi^*}^\mu)\pi^*(\theta).$$

The corresponding primitive cost function $h_R = h_{c_R^*}$ is defined by

$$h_R(P) = \sum_\theta D_{KL}(P_\theta \| P_{\pi^*})\pi^*(\theta).$$

▲

Example 5. Let $\Theta = \{1, \ldots, n\}$. Given $\phi \in Cv(\Delta)$, let $\hat{\phi} : \mathbb{R}_+^n \to \mathbb{R}$ its sublinear extension:

$$\hat{\phi}(z_1, \ldots, z_n) = \begin{cases} \left(\sum_i z_i\right) \phi\left(\frac{z_1}{\sum_i z_i}, \ldots, \frac{z_n}{\sum_i z_i}\right) & \text{if } \sum_i z_i > 0, \\ 0 & \text{otherwise.} \end{cases}$$

The experimental version of c_ϕ is

$$c_\phi^*(\mu) = c_\phi(\mu^*) = \int_\Delta \hat{\phi}\left(\frac{p_1 \pi_1^*}{\bar{\mu}_1}, \ldots, \frac{p_n \pi_n^*}{\bar{\mu}_n}\right) \mathrm{d}\mu(p) - \phi(\pi^*).$$

where, to compute $c_\phi(\mu^*)$, we use the expression for μ^* in Example 3. The corresponding primitive cost function $h_\phi = h_{c_\phi^*}$ is defined by

$$h_\phi(P) = \int_X \hat{\phi}\left(\frac{dP_1}{d\lambda}\pi_1^*, \ldots, \frac{dP_n}{d\lambda}\pi_n^*\right) \mathrm{d}\lambda - \phi(\pi^*).$$

▲

The regularization scheme allows to import tools from rational inattention into the classic framework of statistical decision theory. To illustrate, we go back to the bargaining game from Section 2 and specialize the analysis to the functional form

$$h(P) = k h_R(P) = k \sum_t D_{KL}(P_t \| P_{\sigma^*})\sigma^*(t) \tag{6}$$

where $k > 0$ parametrizes the marginal cost of information and $\sigma^* \in \Delta(T)$ is the uniform distribution over seller's offers. As detailed in Example 4, h_R is dual to c_R^*, the experimental version of the entropic cost c_R. The function c_R is the main specification for the buyer's cost of information in Ravid (2020).

A key tool in applications of rational inattention is the logit representation of optimal information acquisition from Matějka and McKay (2015); we now provide an experimental version of such result. To state it, given a strategy profile (σ, P, β), we denote by β_t the probability that the buyer purchases the good when the seller's offer is t, and by β_σ the average probability that the buyer purchases the good:

$$\beta_t = \sum_x \beta(x) P_t(x) \qquad \text{and} \qquad \beta_\sigma = \sum_t \beta_t \sigma(t).$$

Lemma 5. *Assume (6) for cost of information. If (P, β) is a best response to σ, then*

- *for all $t \in T$,*

$$\beta_t = \frac{e^{\frac{(v-t)\sigma(t)}{k\sigma^*(t)}} \beta_{\sigma^*}}{e^{\frac{(v-t)\sigma(t)}{k\sigma^*(t)}} \beta_{\sigma^*} + 1 - \beta_{\sigma^*}}$$

- *β_{σ^*} is an optimal solution of the maximization problem*

$$\max_{z \in [0,1]} \sum_t \log \left(e^{\frac{(v-t)\sigma(t)}{k\sigma^*(t)}} z + 1 - z \right) \sigma^*(t).$$

Conversely, if z is an optimal solution of the maximization problem above, then there exists a best response (P, β) to σ such that $\beta_{\sigma^*} = z$.

The lemma provides a logit representation of the buyer's best responses. As the proof in the appendix shows, the logit representation generalizes beyond the bargaining game to any information acquisition problem where (6) is the cost of information.

Next, building on lemma 5, we characterize the equilibria of the bargaining game. For simplicity, we focus on the case in which $T = \{2v/3, v, 2v\}$: the seller can offer the good at two-thirds of its value, its value, or twice its value.

The bargaining game admits two classes of equilibria, depending on whether trade occurs or not. In the equilibria where trade does not occur, the seller randomizes between v and $2v$, the buyer runs an uninformative experiment and never purchases the good. The next proposition characterizes the equilibria where trade occurs with positive probability.

Proposition 4. *Assume (6) for cost of information and $T = \{2v/3, v, 2v\}$. If (σ, P, β) is an equilibrium with $\beta_\sigma > 0$, then the following conditions hold:*

- *for all $t \neq v$,*

$$\sigma(t) = \frac{k\sigma^*(t)}{v-t} \log \frac{v(1-\beta_{\sigma^*})}{t - v\beta_{\sigma^*}}.$$

- *for all $t \in T$,*

$$\beta_t = \frac{v}{t}\beta_{\sigma^*}.$$

Conversely, for all $z \in (0, 2/3)$ such that

$$\frac{k}{v}\log\frac{3(1-z)}{2-3z} + \frac{k}{3v}\log\frac{2-z}{1-z} \leq 1,$$

there exists an equilibrium (σ, P, β) with $\beta_\sigma > 0$ and $\beta_{\sigma^*} = z$. Finally, an equilibrium (σ, P, β) with $\beta_\sigma > 0$ exists if and only if

$$k < \frac{3v}{3\log 3 - 2\log 2}.$$

In every equilibrium where trade occurs, the seller is indifferent between all offers: $t\beta_t = t'\beta_{t'}$ for all $t, t' \in T$. The seller randomizes over offers to make it optimal for the buyer to choose a strategy (P, β) such that, indeed, $t\beta_t = t'\beta_{t'}$ for all $t, t' \in T$. In particular, the seller always puts positive probability on $t = 2v/3$ and on $t = 2v$; this creates an incentive for the buyer to acquire information. Finally, there exists an equilibrium where trade occurs if and only if the marginal cost of information k is small relative to the value v of the good; otherwise, the buyer's incentive to acquire information is not strong enough to monitor the seller's offer and she chooses never to buy the good.

5 A taxonomy of cost functions

We study the intersections between the class of experimental cost functions and the main classes of cost functions in rational inattention. We focus on the case in which h is Blackwell monotone and c is monotone in the convex order, the most relevant for applications. Let $\mathcal{H}^{mb} \subseteq \mathcal{H}^b$ the class of cost functions $h : \mathcal{E} \to [0, \infty]$ that are monotone in Blackwell order, $\mathcal{C}^{mx} \subseteq \mathcal{C}^{ex}$ the class of cost functions $c : \Delta^2_+ \to [0, \infty]$ that are monotone in the experimental order, and $\mathcal{C}^{cv} \subseteq \mathcal{C}$ the class of cost functions $c : \Delta^2_+ \to [0, \infty]$ that are monotone in the convex order.

We begin by discussing cost functions that are experimental and convex.

Proposition 5. *(i) If $h \in \mathcal{H}^{mb}$ is convex, so is $c_h \in \mathcal{C}^{mx}$ on each Δ^2_π.*

(ii) If $c \in \mathcal{C}^{mx}$ is convex on $\Delta^2_{\pi^*}$, so is $h_c \in \mathcal{H}^{mb}$.

(iii) If $c \in \mathcal{C}^{cv}$ is convex on $\Delta^2_{\pi^*}$, so is $c^* \in \mathcal{C}^{mx}$ on each Δ^2_π.

Thus convexity is an invariant property of the duality map $h \mapsto c_h$, and it is also preserved by the regularization $c \mapsto c^*$. The result has bite because the mixture operations for random posteriors and experiments can be quite different, as illustrated by the next example.

Example 6. Let $\Theta = \{0, 1\}$. Choose $P, Q \in \mathcal{E}$ such that

$$P_\theta(x) = \begin{cases} 1 & \text{if } \theta = x = 0 \\ 1 & \text{if } \theta = x = 1 \\ 0 & \text{otherwise} \end{cases} \qquad \text{and} \qquad Q_\theta(x) = \begin{cases} 1 & \text{if } \theta = 0 \text{ and } x = 1 \\ 1 & \text{if } \theta = 1 \text{ and } x = 0 \\ 0 & \text{otherwise.} \end{cases}$$

Both experiments P and Q perfectly reveal the state. However, the mixture experiment $P/2 + Q/2$ is completely uninformative:

$$\left(\frac{1}{2}P + \frac{1}{2}Q\right)_\theta (x) = \frac{1}{2}P_\theta(x) + \frac{1}{2}Q_\theta(x) = \begin{cases} 1/2 & \text{if } x \in \{0, 1\} \\ 0 & \text{totherwise.} \end{cases}$$

Thus, for every prior $\pi \in (0, 1)$,

$$\frac{1}{2}B(\pi, P) + \frac{1}{2}B(\pi, Q) \succ_{cv} B\left(\pi, \frac{1}{2}P + \frac{1}{2}Q\right).$$

Mixing the random posteriors $B(\pi, P)$ and $B(\pi, Q)$ and mixing the experiments P and Q lead to quite different results. ▲

The example highlights the difference between mixing in Δ^2 and $\mathcal{E}$. The convex combination $B(\pi, P)/2 + B(\pi, Q)/2$ represents a decision maker with prior π who tosses a fair coin and then decides whether to run experiment P or Q. The convex combination $B(\pi, P/2 + Q/2)$ corresponds to nature privately tossing a fair coin and then reporting to the decision maker the outcome of the experiment P or Q. In this latter case, the decision maker does not know whether experiment P or Q is performed.

The example also shows that Proposition 5 does *not* hold if, in the statement, we replace "convex" with "affine." If the framework is extended to lotteries of experiments, then the induced cost function c_h is convex even if the primitive cost h is not (see, e.g., Ravid, Roesler, and Szentes, 2021).

Next we turn to cost functions that are experimental and lower semicontinuous.

Proposition 6. *(i) If $h \in \mathcal{H}^{mb}$ is lower semicontinuous, so is $c_h \in \mathcal{C}^{mx}$ on Δ^2_+.*

(ii) If $c \in \mathcal{C}^{mx}$ is lower semicontinuous on $\Delta^2_{\pi^*}$, so is $h_c \in \mathcal{H}^{mb}$.

(iii) If $c \in \mathcal{C}^{cv}$ is lower semicontinuous on $\Delta^2_{\pi^*}$, so is $c^* \in \mathcal{C}^{mx}$ on Δ^2_+.

Thus lower semicontinuity is an invariant property of the duality map $h \mapsto c_h$, and it is also preserved by the regularization $c \mapsto c^*$. The result has bite because the notions of convergence for experiments and random posteriors can be quite different, as illustrated by the next example.

Example 7. Let $\Theta = \{0, 1\}$ and $X = [0, 1]$. Let (P_n) be a sequence of experiments such that $P_{n,0}(0) = 1$ and $P_{n,1}(1/n) = 1$. The sequence (P_n) converges to P such that $P_0(0) = P_1(0) = 1$. Every P_n reveals the state perfectly, while P is completely uninformative. Thus, for every prior $\pi \in (0, 1)$,

$$\lim_n B(\pi, P_n) = (1-\pi)\delta_0 + \pi\delta_1 \succ_{cv} \delta_\pi = B(\pi, P).$$

The map $P \mapsto B(\pi, P)$ is not continuous. ▲

The example highlights the difference between the topologies Δ^2 and $\mathcal{E}$.[14] In particular, the example shows that Proposition 6 does not hold if, in the statement, we replace "lower semicontinuous" with "continuous."

Proposition 6 also shows that a cost function $c \in \mathcal{C}^{mx}$ is lower semicontinuous on the *subdomain* $\Delta^2_{\pi^*}$ if and only if it is lower semicontinuous on the *entire domain* Δ^2_+. Thus, for experimental cost functions, "local" lower semicontinuity is equivalent to "global" lower semicontinuity.

Following Caplin and Dean (2015) and De Oliveira, Denti, Mihm, and Ozbek (2017), we introduce a class of cost functions.

Definition 6. A cost $c \in \mathcal{C}^{cv}$ is *canonical* if it is convex and lower semicontinuous on each Δ^2_π.

The majority of cost functions considered in applications are canonical. From Propositions 5 and 6, it follows immediately that

- if $h \in \mathcal{H}^{mb}$ is convex and lower semicontinuous, then $c_h \in \mathcal{C}^{mx}$ is canonical;
- if $c \in \mathcal{C}^{mx}$ is canonical, then $h_c \in \mathcal{H}^{mb}$ is convex and lower semicontinuous;
- if $c \in \mathcal{C}^{cv}$ is canonical, so is $c^* \in \mathcal{C}^{mx}$.

These observations motivate the following definition:

Definition 7. A cost $h \in \mathcal{H}^{mb}$ is *canonical* if it is convex and lower semicontinuous.

It follows from Theorem 2 of De Oliveira, Denti, Mihm, and Ozbek (2017) that every canonical $c \in \mathcal{C}^{cv}$ admits a variational representation:

[14] Other topologies on $\mathcal{E}$ have been considered in the literature; see Section **??** for discussion.

Lemma 6. *A cost $c \in \mathcal{C}^{cv}$ is canonical if and only if for every $\pi \in \Delta_+$ there is a set $\Phi^\pi \subseteq Cv(\Delta)$, with $\sup_{\phi \in \Phi^\pi} \phi(\pi) < \infty$, such that*

$$c(\mu) = \sup_{\phi \in \Phi^\pi} \int_\Delta \phi(p) \, \mathrm{d}\mu(p) - \sup_{\phi \in \Phi^\pi} \phi(\pi) \qquad \forall \mu \in \Delta^2_\pi.$$

Building on Lemma 6, the next result provides variational representations for the canonical elements of $\mathcal{C}^{mx}$ and $\mathcal{H}^{mb}$. In what follows, with a slight abuse of notation we denote by 1 the vector of ones $(1, \ldots, 1) \in \mathbb{R}^\Theta$. To shorten notation, we also omit the restriction $\sup_{\phi \in \Phi^\pi} \phi(\pi) < \infty$. We adopt instead the following convention: if $\sup_{\phi \in \Phi^\pi} \phi(\pi) = \infty$, then $c(\mu) = 0$ for $\mu = \delta_\pi$, and $c(\mu) = \infty$ for $\mu \succ_{cv} \delta_\pi$.

Proposition 7. *(i) A cost $c \in \mathcal{C}^{mx}$ is canonical if and only if there is a set $\Psi \subseteq Cs(\mathbb{R}^\Theta_+)$ such that*

$$c(\mu) = \sup_{\psi \in \Psi} \int_\Delta \psi\left(\frac{p}{\bar{\mu}}\right) \mathrm{d}\mu(p) - \sup_{\psi \in \Psi} \psi(1) \qquad \forall \mu \in \Delta^2_+.$$

(ii) A cost $h \in \mathcal{H}^{mb}$ is canonical if and only if there is a set $\Psi \subseteq Cs(\mathbb{R}^\Theta_+)$ such that

$$h(P) = \sup_{\psi \in \Psi} \int_{\mathbb{R}^\Theta_+} \psi\left(\frac{\mathrm{d}P}{\mathrm{d}\lambda}\right) \mathrm{d}\lambda - \sup_{\psi \in \Psi} \psi(1) \qquad \forall P \in \mathcal{E}.$$

Following Caplin and Dean (2013), we define a subclass of canonical cost functions.

Definition 8. A cost $c : \Delta^2_+ \to [0, \infty]$ is *posterior separable* if for every $\pi \in \Delta_+$ there is a convex and lower semicontinuous function $\phi^\pi : \Delta \to (-\infty, \infty]$ such that

$$c(\mu) = \int_\Delta \phi^\pi(p) \, \mathrm{d}\mu(p) - \phi^\pi(\pi) \qquad \forall \mu \in \Delta^2_\pi.$$

A chief example of posterior separability is the entropy cost c_R, defined in (4), which corresponds to the integrand

$$\phi^\pi(p) = \sum_\theta p(\theta) \log p(\theta).$$

For the entropic cost function, $\phi^\pi(p)$ is a finite quantity for all $p \in \Delta$. The cost function proposed by Morris and Strack (2019) is an example of a posterior separable cost function whose integrand can be infinite:

$$\phi^\pi(p) = \begin{cases} \sum_{\theta,\tau} p(\theta) \log \frac{p(\theta)}{p(\tau)} & \text{if } p \in \Delta_+ \\ \infty & \text{otherwise.} \end{cases}$$

Both cost functions are representatives of the subclass of *uniformly* posterior separable cost functions, for which ϕ^π is independent of π.

We have already discussed the inconsistency between experimental cost functions and uniform posterior separability (see Corollary 1). Next we characterize the cost functions that are experimental and posterior separable.

Proposition 8. *A cost $c \in \mathcal{C}^{mx}$ is posterior separable if and only if there is a sublinear and lower semicontinuous function $\psi : \mathbb{R}_+^\Theta \to (-\infty, \infty]$ such that*

$$c(\mu) = \int_\Delta \psi\left(\frac{p}{\bar{\mu}}\right) \mathrm{d}\mu(p) - \psi(1) \qquad \forall \mu \in \Delta_+^2.$$

The result suggests the following definition:

Definition 9. A cost $h \in \mathcal{H}^{mb}$ is *likelihood separable* if there is a sublinear and lower semicontinuous function $\psi : \mathbb{R}_+^\Theta \to (-\infty, \infty]$ such that

$$h(P) = \int_{\mathbb{R}_+^\Theta} \psi\left(\frac{\mathrm{d}P}{\mathrm{d}\lambda}\right) \mathrm{d}\lambda - \psi(1) \qquad \forall P \in \mathcal{E}.$$

Posterior separability and likelihood separability are dual notions:

Proposition 9. *(i) If $h \in \mathcal{H}^{mb}$ is likelihood separable, then $c_h \in \mathcal{C}^{mx}$ is posterior separable.*
(ii) If $c \in \mathcal{C}^{mx}$ is posterior separable, then $h_c \in \mathcal{H}^{mb}$ is likelihood separable.
(iii) If $c \in \mathcal{C}^{cv}$ is posterior separable, so is $c^* \in \mathcal{C}^{mx}$.

We can illustrate the result with the following version of the duality diagram:

$$\begin{array}{ccccccc}
 & h & \xrightarrow{D} & c_h & & c^* \longleftarrow c & \\
\mathcal{H}_{ls}^{mb} & & & & \mathcal{C}_{ps}^{mx} & & \mathcal{C}_{ps}^{cv} \\
 & h_c & \xleftarrow{D^{-1}} & c & & &
\end{array}$$

where the subscripts "ls" and "ps" stand for likelihood and posterior separable. In particular, the duality implies that h_R—obtained from the regularization of rational inattention, see Example (4) for the definition—is likelihood separable.

Functionals of experiments that are, in our language, likelihood separable are central to statistical decision theory. To refer to the elements of $\mathcal{H}_{ls}^{mb}$, Torgersen (1991, pp. 351-360) speaks of "monotone representable" functionals of experiments.[15] When the state is binary, likelihood separable cost functions correspond one-to-one to f-divergences, a major class of statistical distances (Csiszár, 1963; Ali and Silvey, 1966).

[15]We prefer "likelihood separable" to "monotone representable" in order to emphasize the duality with "posterior separable," a denomination that is by now established in economics.

Definition 10. Let $f : (0, \infty) \to \mathbb{R}$ be a convex function. The *f-divergence* between $\xi_1, \xi_2 \in \Delta(X)$ is the quantity

$$D_f(\xi_1 \| \xi_2) = \int_X \frac{\mathrm{d}\xi_2}{\mathrm{d}\lambda} f\left(\frac{\mathrm{d}\xi_1/\mathrm{d}\lambda}{\mathrm{d}\xi_2/\mathrm{d}\lambda}\right) \mathrm{d}\lambda$$

where λ is control measure for ξ_1 and ξ_2.[16]

Proposition 10. *Let $\Theta = \{0, 1\}$. A cost function $h : \mathcal{E} \to [0, \infty]$ is likelihood separable, with ψ finite on the interior of its domain, if and only if there is a convex function $f : (0, \infty) \to \mathbb{R}$ such that*

$$h(P) = D_f(P_0 \| P_1) - f(1) \qquad \forall P \in \mathcal{E}.$$

Thus f-divergences can be used to construct likelihood separable cost functions.[17] A main example of f-divergence is the Kullback-Leibler divergence, which corresponds to $f(t) = t \log t$ (see Example 4). Proposition 10 implies that the cost

$$h(P) = D_{KL}(P_0 \| P_1)$$

is likelihood separable. The cost function proposed by Pomatto, Strack, and Tamuz (2020) can be seen as a generalization beyond the binary case.

6 Information acquisition problems

We apply the duality between $h \in \mathcal{H}^b$ and $c \in \mathcal{C}^{ex}$ to information acquisition problems.

6.1 Primal form

Consider a decision maker who must choose among a set of alternative actions whose consequences depend on the state of nature, which is uncertain. Formally, there is a finite set A of available *actions* a that can result in different material *consequences* c, within a set C, depending on which state $\theta \in \Theta$ obtains. In particular, the dependence of consequences on actions and states is described by a *consequence function* $\zeta : A \times \Theta \to C$. Finally, a *utility function* $u : C \to \mathbb{R}$ ranks consequences. The quintet

$$(A, \Theta, C, \zeta, u)$$

is a *decision problem under uncertainty*.

[16] Behind the integral is adopted the conventions $f(0) = \lim_{t \to 0} f(t)$ and $0 \cdot f(t/0) = t \lim_{s \to \infty} f(s)/s$.
[17] Liese and Vajda (2006) present related results on f-divergences and measures of information.

Suppose the decision maker can choose to perform an experiment $P \in \mathcal{E}$ that (probabilistically) generate signals $x \in X$ that give information about the state. The septet

$$(A, \Theta, C, X, \mathcal{E}, \zeta, u)$$

is an *information acquisition problem.* Here the function $\zeta : A \times \Theta \times \mathcal{E} \to C$ details, for each experiment P, a consequence $\zeta(a, \theta, P)$ determined by selecting action a when state θ obtains.[18] Information acquisition problems were first formulated by Wald (1950). Wald was mostly interested in representing statisticians as decision makers, so he speaks of *statistical decision problems.*

As it is often convenient, we distinguish between "gross" consequences and costs. We assume that each experiment P has a monetary cost $h(P) \in [0, \infty]$, for instance determined by its market price. Taking action a in state θ leads to a gross consequence $\gamma(a, \theta)$, within a set G. The "net" consequence $\zeta(a, \theta, P)$ can be decomposed into gross consequence and cost. Specifically, we assume

$$\gamma(a, \theta) = \gamma(b, \tau) \text{ and } h(P) = h(Q) \quad \Longrightarrow \quad \zeta(a, \theta, P) = \zeta(b, \tau, Q). \tag{7}$$

The next example illustrates two main cases.

Example 8. (i) The gross consequence is monetary as well. Here $C \subseteq [-\infty, \infty)$ and

$$\zeta(a, \theta, P) = \gamma(a, \theta) - h(P).$$

The utility function u is over monetary amounts, with

$$(u \circ \zeta)(a, \theta, P) = u(\gamma(a, \theta) - h(P)).$$

The formulation arises in applications to macroeconomics and finance, e.g., the decision maker is an investor who can acquire additional information about asset returns before re-balancing her portfolio (Cabrales, Gossner, and Serrano, 2013).

(ii) The gross consequence is not monetary. Here $C = G \times [0, \infty]$ and

$$\zeta(a, \theta, P) = (\gamma(a, \theta), h(P)).$$

Via a quasi-linear utility function $u : G \times [0, \infty] \to [-\infty, \infty)$ given by $u(g, t) = u_0(g) - t$,

[18] In certain applications, the material consequence of the decision makers' behavior could depend also on the realized signal; if so, the domain of the consequence function should be $A \times \Theta \times \mathcal{E} \times X$.

with $u_0(g) = u(g, 0)$, we then have

$$(u \circ \zeta)(a, \theta, P) = u_0(\gamma(a, \theta)) - h(P).$$

Here u_0 provides a representation of gross consequences in monetary terms. The quasi-linear formulation is common in mechanism design (Persico, 2000; Bergemann and Välimäki, 2002) and discrete choice (Matějka and McKay, 2015). ▲

The decision maker is Bayesian with prior belief $\pi \in \Delta_+$.[19] Solving the information acquisition problem, the decision maker first selects and performs an experiment, then updates her beliefs as a result of the signal she receives, and finally takes an action that is optimal given the posterior belief. We can formalize the information acquisition problem as follows:

$$\max_{P \in \mathcal{E}} \int_X \left(\max_{a \in A} \sum_\theta v(a, \theta, h(P)) p_x(\theta) \right) \mathrm{d}P_\pi(x) \tag{8}$$

where $v : A \times \Theta \times [0, \infty] \to [-\infty, \infty)$ is the *payoff function* given by

$$v(a, \theta, t) = u(\zeta(a, \theta, P)) \qquad \text{with} \qquad t = h(P).$$

The payoff function is well defined by (7).

Example 8 (Continued)**.** In the monetary case, the information acquisition problem is

$$\max_{P \in \mathcal{E}} \int_X \left(\max_{a \in A} \sum_\theta u(\gamma(a, \theta) - h(P)) p_x(\theta) \right) \mathrm{d}P_\pi(x),$$

while in the quasi-linear case it becomes

$$\max_{P \in \mathcal{E}} \int_X \left(\max_{a \in A} \sum_\theta v_0(a, \theta) p_x(\theta) \right) \mathrm{d}P_\pi(x) - h(P)$$

where $v_0(a, \theta) = u_0(\gamma(a, \theta))$ is the gross payoff from action a in state θ.

The quasi-linear specification has a convenient separable form. In the monetary case, such a form arises when u is CARA and we move to certainty equivalents:

$$\max_{P \in \mathcal{E}} u^{-1} \left(\int_X \left(\max_{a \in A} \sum_\theta u(\gamma(a, \theta)) p_x(\theta) \right) \mathrm{d}P_\pi(x) \right) - h(P).$$

The result follows from some simple algebra. ▲

[19] Wald also considers non-Bayesian decision makers—most notably, decision makers who evaluate alternatives on the basis of the worst case scenario.

Let $\boldsymbol{A}$ be the set of action rules $\boldsymbol{a} : \mathcal{E} \times X \to A$ such that, for every $P \in \mathcal{E}$, the section $\boldsymbol{a}_P : X \to A$ is measurable. A *solution* of the information acquisition problem is a pair $(\hat{\boldsymbol{a}}, \hat{P}) \in \boldsymbol{A} \times \mathcal{E}$ such that (i) each action $\hat{\boldsymbol{a}}_P(x)$ is optimal given each experiment P and signal x, that is, for all $a \in A$

$$\sum_\theta v\left(\hat{\boldsymbol{a}}_P(x), \theta, h(P)\right) p_x(\theta) \geq \sum_\theta v\left(a, \theta, h(P)\right) p_x(\theta),$$

and (ii) $\hat{P}$ is optimal given the action rule $\hat{\boldsymbol{a}}_{\hat{P}}$ that it determines, that is, for all $P \in \mathcal{E}$,

$$\sum_\theta \pi(\theta) \int_X v(\hat{\boldsymbol{a}}_{\hat{P}}(x), \theta, h(\hat{P}))\, \mathrm{d}\hat{P}_\theta(x) \geq \sum_\theta \pi(\theta) \int_X v\left(\hat{\boldsymbol{a}}_P(x), \theta, h(P)\right) \mathrm{d}P_\theta(x).$$

6.2 Dual form

The duality established by Theorem 1 between $\mathcal{H}^b$ and $\mathcal{C}^{ex}$ leads to a dual version of (8). Given a cost $c : \Delta_+^2 \to [0, \infty]$, the *dual information acquisition problem* is

$$\max_{\mu \in \Delta_\pi^2} \int_\Delta \left(\max_{a \in A} \sum_\theta v(a, \theta, c(\mu)) p(\theta) \right) \mathrm{d}\mu(p). \tag{9}$$

To relate (8) and (9), we define $V : \Delta_+ \times \mathcal{E} \to [-\infty, \infty)$ and $W : \Delta_+^2 \to [-\infty, \infty)$ by

$$V(\pi, P) = \int_X \left(\max_{a \in A} \sum_\theta v(a, \theta, h(P)) p_x(\theta) \right) \mathrm{d}P_\pi(x)$$

and

$$W(\mu) = \int_\Delta \left(\max_{a \in A} \sum_\theta v(a, \theta, c(\mu)) p(\theta) \right) \mathrm{d}\mu(p).$$

Proposition 11. *For $h \in \mathcal{H}^b$ and $c \in \mathcal{C}^{ex}$ such that $c = c_h$, we have:*

(i) $\max_{P \in \mathcal{E}} V(P, \pi) = \max_{\mu \in \Delta_\pi^2} W(\mu)$*;*

(ii) if $\hat{P}$ maximizes $V(\pi, P)$, then $B(\pi, \hat{P})$ maximizes $W(\mu)$ for $\mu \in \Delta_\pi^2$;

(iii) if $\hat{\mu}$ maximizes $W(\mu)$ for $\mu \in \Delta_\pi^2$, then $P^{\hat{\mu}}$ maximizes $V(\pi, P)$.

We can diagram the duality between solutions as follows:

$$\hat{P} \quad \overset{B_\pi}{\underset{B_\pi^{-1}}{\rightleftarrows}} \quad \hat{\mu}$$

Depending on the context, it may be convenient to study information acquisition in primal or in dual form. The next example provides an illustration for quasi-linear environments.

Example 9. In quasi-linear environments, the primal information acquisition problem is

$$\max_{P\in\mathcal{E}}\int_X\left(\max_{a\in A}\sum_\theta v_0(a,\theta)p_x(\theta)\right)\mathrm{d}P_\pi(x)-h(P). \tag{10}$$

When h is canonical (see Definition 7), we can conveniently reduce (10) to a finite-dimensional convex program. To illustrate, let $\mathcal{E}_A$ be the set of experiments $P:\Theta\to\Delta(A)$ whose signal space is the set of actions. Being h Blackwell monotone, from standard arguments in the stile of the revelation principle it follows that (10) is equivalent to

$$\max_{P\in\mathcal{E}_A}\sum_{\theta,a}v_0(a,\theta)P_\theta(a)\pi(\theta)-h(P). \tag{11}$$

If $(\hat{\boldsymbol{a}},\hat{P})$ is a solution of (10), then $\hat{Q}_\theta(a)=\hat{P}_\theta(\hat{\boldsymbol{a}}_{\hat{P}}=a)$ defines a solution of (11); if $\hat{Q}$ is a solution of (11), there there is a solution $(\hat{\boldsymbol{a}},\hat{P})$ of (10) such that $\hat{Q}_\theta(a)=\hat{P}_\theta(\hat{\boldsymbol{a}}_{\hat{P}}=a)$.

When h is canonical, (11) is a finite-dimensional convex program. A canonical cost function h is convex and lower semicontinuous. In addition, being Θ and A finite, the set $\mathcal{E}_A$ can be identified with a convex subset of a finite-dimensional Euclidean space. Thus (11) can be studied as a finite-dimensional convex program (e.g., with Lagrange multipliers).

When h is likelihood separable (see Definition 9), it may be convenient to study information acquisition in dual form, for it reduces to a concavification. The dual information acquisition problem is

$$\max_{\mu\in\Delta^2_\pi}\int_\Delta\left(\max_{a\in A}\sum_\theta v_0(a,\theta)p(\theta)\right)\mathrm{d}\mu(p)-c_h(\mu). \tag{12}$$

When h is likelihood separable with integrand $\psi:\mathbb{R}^\Theta_+\to(-\infty,\infty]$, the dual cost function c_h is posterior separable with integrands $\phi^\pi:\Delta\to(-\infty,\infty]$ defined by $\phi^\pi(p)=\psi(p/\pi)$ (see Proposition 9). As observed by Caplin and Dean (2013), when the cost function is posterior separable, (12) reduces to a concavification:

$$\max_{\mu\in\Delta^2_\pi}\int_\Delta\left(\max_{a\in A}\sum_\theta v_0(a,\theta)p(\theta)-\psi\left(\frac{p}{\pi}\right)\right)\mathrm{d}\mu(p)+\psi(1).$$

Concavifications feature prominently in economics, e.g., in Bayesian persuasion. ▲

7 Sequential information acquisition

We consider an extension of our framework where information acquisition is sequential.

7.1 Setup and analysis

Let X^n be the Cartesian product of $n \in \mathbb{N}$ copies of the signal space X; we endow X^n with the product topology and the corresponding Borel σ-algebra.

Suppose the agent can perform a finite sequence of experiments $P^n = (P_1, \ldots, P_n)$. The i-th experiment in the sequence may depend on the outcomes of the first $i-1$ experiments, so we represent it as a Borel measurable function $P_i : X^{i-1} \times \Theta \to \Delta(X)$. We adopt the convention that $X^0 = \{\varnothing\}$.

We denote by $\mathcal{E}^n$ the set of sequential experiments of length n; clearly, $\mathcal{E}^1 = \mathcal{E}$. Let $P_{x^{i-1}} \in \mathcal{E}$ be the one-shot experiment that the decision maker performs after an history of signals $x^{i-1} \in X^{i-1}$. We focus on a basic specification where there is no discounting and the flow cost of experimentation

$$h\left(P_{x^{i-1}}\right) \in [0, \infty] \tag{13}$$

is independent of the previous experiments $P_1, \ldots, P_{i-1}$ and their outcomes $x_1, \ldots, x_{i-1}$.

Given a prior belief $\pi \in \Delta_+$, the expected cost of the sequential experiment P^n is

$$h(\pi, P^n) = \sum_{i=1}^{n} \int_{X^{i-1}} h\left(P_{x^{i-1}}\right) \mathrm{d}P_\pi^{i-1}(x^{i-1})$$

where $P_\pi^{i-1} \in \Delta(X^{i-1})$ is the predictive probability generated by the prior π and the first $i-1$ experiments $P_1, \ldots, P_{i-1}$. For the integral to be well defined, we assume that the flow cost function $h : \mathcal{E} \to [0, \infty]$ is Borel measurable.

The induced cost $c_{h^n} : \Delta_+^2 \to [0, \infty]$ over random posteriors is

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

where $B(\bar{\mu}, P^n)$ is the random posterior generated by prior $\bar{\mu}$ and sequential experiment P^n. The random posterior $B(\bar{\mu}, P^n)$ is the distribution of the posterior beliefs that the decision maker would hold after observing the full sequence of signals $x^n = (x_1, \ldots, x_n)$ and updating her prior belief π accordingly.

The cost function c_{h^n} corresponds to the case in which the agent can perform any sequential experiment of length n. We write $c_{h^\infty} : \Delta_+^2 \to [0, \infty]$ for the case in which the decision maker can perform any sequential experiment of any length:

$$c_{h^\infty}(\mu) = \inf_n c_{h^n}(\mu).$$

Definition 11. A cost function $c : \Delta^2_+ \to [0, \infty]$ is *experimental of order* $n = 1, 2, \ldots, \infty$ if there is a flow cost function $h : \mathcal{E} \to [0, \infty]$ such that $c = c_{h^n}$.

An experimental cost function of order $n = 1$ is simply an experimental cost function, as we defined in Section 4. In Theorem 1 we showed that a cost function is experimental if and only if it is invariant under $\sim_{ex}$, the symmetric part of the experimental order. When sequential experiments are allowed, the induced cost function on random posteriors may not be invariant under $\sim_{ex}$, as the next example shows.

Example 10. Let $\Theta = \{0, 1\}$. Take a one-shot experiment P given by

$$P_\theta(x) = \begin{cases} 1 & \text{if } \theta = 0 \text{ and } x = 0 \\ 0 & \text{if } \theta = 0 \text{ and } x = 1 \\ \frac{1}{2} & \text{if } \theta = 1 \text{ and } x \in \{0, 1\}. \end{cases}$$

The experiment P has two possible outcomes, 0 and 1. If $\theta = 0$, then $x = 0$ with probability one. If $\theta = 1$, then $x = 0$ and $x = 1$ are equally likely.

Consider a primitive cost function $h : \mathcal{E} \to [0, \infty)$ given by

$$h(Q) = \begin{cases} 0 & \text{if } P \succeq_b Q \\ 1 & \text{otherwise.} \end{cases}$$

Every experiment that is less informative than P is free. Every other experiment costs one.

Suppose that the decision maker can perform sequential experiments of length $n = 2$. Starting from a prior $\pi \in (0, 1)$, the least expensive way to perfectly learn the state and obtain random posterior $(1 - \pi)\delta_0 + \pi\delta_1$ is

- run experiment P first;
- if the outcome is $x = 1$, stop experimenting;
- if the outcome is $x = 0$, run an experiment that perfectly reveals the state.

The expected cost of the procedure is the probability of running the second experiment:

$$c_{h^2}\left((1 - \pi)\delta_0 + \pi\delta_1\right) = 1 - \frac{\pi}{2}.$$

If $\rho \in (0,1)$, then $(1-\rho)\delta_0 + \rho\delta_1 \sim_{ex} (1-\pi)\delta_0 + \pi\delta_1$. However, $c_{h^2}((1-\pi)\delta_0 + \pi\delta_1) = c_{h^2}((1-\rho)\delta_0 + \rho\delta_1)$ if and only if $\pi = \rho$. Thus c_{h^2} is not invariant under $\sim_{ex}$. ▲

As the example shows, when the decision maker can perform sequential experiments, the induced cost function on random posteriors may not be invariant under $\sim_{ex}$. Yet, not *any* cost function on random posteriors can be induced:

Proposition 12. *Let $\Theta = \{0,1\}$. If $c : \Delta^2_+ \to [0,\infty]$ is lower semicontinuous on each subdomain Δ^2_π, monotone in the convex order, and experimental of order $n = 1,2,\ldots,\infty$, then the following conditions are equivalent:*

(i) For some $\theta \in \Theta$, $\lim_{\bar{\mu}(\theta)\to 1} c(\mu) = 0$.

(ii) For all $\mu \in \Delta^2_+$, $c(\mu) = 0$.

For the benchmark case of a binary state space, Proposition 12 extends Corollary 1. In particular, it implies that a cost function is experimental of order $n = 1,2,\ldots,\infty$, uniformly posterior separable, and bounded if and only if it is identical to zero. For example, the entropic cost function used by Matějka and McKay (2015) is uniformly posterior separable, bounded, and not identical to zero; thus it is not experimental of any order n. Thus, the inconsistency between rational inattention and a primitive model of costly experimentation holds regardless of whether experimentation is one-shot or sequential.[20]

As attested by Proposition 12, the experimental approach puts discipline on the relation between prior beliefs and cost of information, regardless of whether the decision maker can perform one-shot or sequential experiments. Next we provide another manifestation.

Proposition 13. *If $c : \Delta^2_+ \to [0,\infty]$ is experimental of order $n = 1,2,\ldots,\infty$, then for all $\pi, \rho \in \Delta_+$, $\alpha \in [0,1]$, and $P \in \mathcal{E}$*

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

The result motivates the following definition:

Definition 12. A cost function $c : \Delta^2_+ \to [0,\infty]$ is *concave in the prior* if for all $\pi, \rho \in \Delta_+$, $\alpha \in [0,1]$, and $P \in \mathcal{E}$

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

A cost function can be concave in the prior but not experimental; for example, the entropic cost function c_R is concave in the prior (see Cover and Thomas, 2012, Theorem

[20] One can derive stringent restrictions even without assuming that c is lower semicontinuous and monotone in the convex order; see Claim 4 in the appendix.

2.7.4) but not experimental of any order n (Proposition 12). Next we provide an example from rational inattention of a cost function that is not concave in the prior.

Example 11. Many applications of rational inattention adopts the specification $c_K : \Delta^2_+ \to [0, \infty]$ given by

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

where $K > 0$ is interpreted as a bound on the capacity of the decision maker to process information. The cost function c_K is *quasi-concave* but *not concave* in the prior. It is a monotone transformation of c_R, which is concave—and therefore quasi-concave—in the prior. As well known, monotone transformations preserve quasi-concavity but not concavity: it is easy to come up with instances of $\pi, \rho \in \Delta_+$, $\alpha \in [0, 1]$, and $P \in \mathcal{E}$ such that

$$c_K(B(\alpha\pi + (1-\alpha)\rho, P)) < \alpha c_K(B(\pi, P)) + (1-\alpha)c_K(B(\rho, P)).$$

▲

In a context unrelated to sequential experiments, Miao and Xing (2020) discuss the relation between cost of information and concavity in the prior. Assuming uniform posterior separability, they show that, when the cost of information is concave in the prior, the value function in information acquisition problems is convex in the prior. Next we generalize their result beyond uniform posterior separability.

Proposition 14. *Let $c : \Delta^2_+ \to [0, \infty]$ be a cost function defined on random posteriors. Consider the information acquisition problem*

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

where π is a prior belief, A is a finite set of actions, and $v : A \times \Theta \to \mathbb{R}$ is a utility function. If c is concave in the prior, then $V(\pi)$ is a convex function of $\pi \in \Delta_+$.

The convexity in the prior of the value function—and therefore the concavity in the prior of the cost 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 the information today.

Propositions 12 and 13 provide necessary conditions for a cost function to be experimental of order $n = 1, 2, \dots, \infty$. Our analysis leaves open the question of what conditions are necessary and sufficient.

7.2 Discussion

The analysis of sequential information acquisition was pioneered by Wald (1947) and Arrow, Blackwell, and Girshick (1949).

A few recent papers have explored the relation between sequential information acquisition and rational inattention; the closest to ours are Hébert and Woodford (2019), Morris and Strack (2019), and Bloedel and Zhong (2020). These papers address questions similar to ours, but allow the flow cost (13) to depend arbitrarily on the evolving beliefs of the decision maker, as in rational inattention. As a consequence, more cost functions on random posteriors can be generated. The next example from Bloedel and Zhong (2020) provides a concrete illustration.

Example 12 (Bloedel and Zhong, 2020)**.** Let $\phi \in Cv(\Delta)$. In contrast with (13), suppose that the flow cost of running experiment $P_{x^{i-1}}$ is given by

$$c_\phi \left(B \left(p_{x^{i-1}}, P_{x^{i-1}}\right)\right) = \int_\Delta \phi \, \mathrm{d}B \left(p_{x^{i-1}}, P_{x^{i-1}}\right) - \phi \left(p_{x^{i-1}}\right)$$

where $p_{x^{i-1}} \in \Delta$ is the belief that the decision maker holds after a history of signals $x^{i-1} = (x_1, \dots, x_{i-1})$. Given prior $\pi \in \Delta_+$, the resulting cost of the sequential experiments P^n is

$$h(\pi, P^n) = \sum_{i=1}^n \int_{X^{i-1}} c_\phi \left(B \left(p_{x^{i-1}}, P_{x^{i-1}}\right)\right) \, \mathrm{d}P_\pi^{i-1}(x^{i-1}).$$

Note, in particular, that $h(\pi, P^n)$ is not an affine function of π. As Bloedel and Zhong show, $c_\phi(B(\pi, P^n)) = h(\pi, P^n)$. Thus, for all $\mu \in \Delta^2_+$,

$$c_\phi(\mu) = \inf\{h(\pi, P^n) : B(\bar{\mu}, P^n) = \mu\}.$$

▲

Thus, if the flow cost depends arbitrarily on the evolving beliefs of the decision maker, then any uniformly posterior separable cost function can be generated—e.g., the entropy cost of Matějka and McKay (2015). Our Proposition 12 adds a caveat: the arbitrariness is crucial; if the flow cost depends only on the per-period experiment, then no cost function that is uniformly posterior separable and bounded can be generated.

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 can arbitrarily depend on the decision maker's evolving beliefs, as in Example 12, the exercise loses some of its appeal. In a circular fashion, the problem of explaining the dependence of the ex-ante 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.

More concretely, consider again the bargaining game of Section 2, our motivating example. Suppose that the buyer can perform multiple experiments in sequence; for example, suppose that the buyer can perform any sequential experiment of any length. By Example 12, the results of Ravid (2020), here detailed by Propositions 1 and 2, extend to the case of sequential experiments. It is easy to see that our Proposition 3 also extends to the case of sequential experiments. Thus, in games with information acquisition, the rational inattention model and the experimental approach we propose are substantially different, regardless of whether players can perform one-shot or sequential experiments.

As a special case of their framework, Bloedel and Zhong (2020) study what happens when the flow cost depends only on the per-period experiment, as in (13). Their findings are broadly consistent with ours. In particular, under the hypothesis that h is locally quadratic, they show that no bounded, non-trivial, uniformly posterior separable cost function is consistent with a primitive model of sequential information acquisition (Bloedel and Zhong, 2020, Proposition 3).[21] Our Proposition 12 complements their result, as we put no functional form assumption on h, generalize beyond uniform posterior separability, and rely on a different argument for the proof. In their appendix, Bloedel and Zhong also point out that concavity in the prior is a necessary condition for a cost function on random posteriors to be consistent with a primitive model of sequential information acquisition.

Our analysis leaves open the questions of what cost functions are experimental of order $n \geq 2$. Morris and Strack (2019) provide an answer in a specific environment. They consider a binary state space $\Theta = \{0, 1\}$. The decision maker observes the evolution of a Brownian motion whose drift depends on the state. The flow cost is a function of the passage of time, which is of course independent of the evolving beliefs of the decision maker. Morris and Strack characterize the induced cost function on random posteriors: given $\pi \in (0, 1)$,

$$c_{MS}(B(\pi, P)) = (1 - \pi) D_{KL}(P_0 \| P_1) + \pi D_{KL}(P_1 \| P_0).$$

Their result is consistent with our findings. First, the hypothesis of Proposition 12 is not

[21] When there are al least three states, their result holds for all non-trivial uniformly posterior separable cost functions, bounded or unbounded.

satisfied. For all $\pi \in (0,1)$, $c_{MS}((1-\pi)\delta_0 + \pi\delta_1) = \infty$. Thus

$$\lim_{\pi \to 1} c_{MS}((1-\pi)\delta_0 + \pi\delta_1) = \infty > 0.$$

Moreover, $c_{MS}(B(\pi, P))$ is affine in π, so, in particular, concave in π.

In Morris and Strack (2019), the arrival of information follows a continuous-time process, Brownian motion. Bloedel and Zhong (2020) re-derive Morris and Strack's result in a discrete time model with flexible information acquisition. In the language of our paper, they consider a binary state space $\Theta = \{0,1\}$ and a primitive cost function $h : \mathcal{E} \to [0, \infty]$ given by

$$h(P) = \max\{D_{KL}(P_0 \| P_1), D_{KL}(P_0 \| P_1)\}.$$

They show that $c_{MS} = c_{h^\infty}$ (Bloedel and Zhong, 2020, Proposition 3). To date, that is the only non-trivial, complete solution of the problem we discuss in this section.

8 Concluding remarks

Producing vs processing information. The literature distinguishes between two main activities of information acquisition: information production and information processing. Borrowing an example from Sims (2010, p. 161), finding whether a well test indicates oil is an instance of information production; reading a report on the well test is an instance of information processing. These two activities are obviously very different, but we believe that our framework is rich enough to represent both.

In the case of information production, the state $\theta \in \{oil, not\}$ represents whether there is oil in the field or not. An experiment P represents a particular choice of well test, with associated monetary cost $h(P)$. In the case of information processing, the state $\theta \in \{yes, no\}$ represents whether it is reported that there is oil in the field or not. An experiment P represents a level of attention in reading the report, with associated psychological cost $h(P)$. In the case of information production, information acquisition is a tangible activity with a pecuniary cost. In the case of information processing, information acquisition is an "as if" story to model limited attention.

Ultimately, it is an empirical question whether the framework we propose is a good representation of these two activities. On a theoretical level, however, we claim that there are advantages in the experimental approach we propose, regardless of whether the decision maker produces or processes information. A proof-of-concept is information acquisition in games. In Ravid (2020), the buyer's cost of information can have both pecuniary components (e.g., consulting fees) and psychological components (e.g., mental effort to understand a complex contract). As we show in our leading application, the experimental approach can substantially simplify the analysis of the buyer's information acquisition problem and of the strategic interaction with the seller's incentives.

Information and non-common priors. The normalization map $\mu \mapsto \mu^*$ is a central object of our analysis. It admits the following multi-agent interpretation: Consider two agents with full support, non-common priors π and π^*. Given a common experiment P, the agent with prior π ends up with random posterior $\mu = B(\pi, P)$ and the agent with prior π^* ends up with random posterior $\mu^* = B(\pi^*, P)$. As we detail in Example 3, given signal x, the posterior beliefs p_x and p_x^* are related by

$$p_x^*(\theta) = \frac{p_x(\theta)(\pi^*(\theta)/\pi(\theta))}{\sum_\tau p_x(\tau)(\pi^*(\tau)/\pi(\tau))}.$$

The same relation pops up in many papers on information economics with multiple agents and non-common priors. Most recently, the relation has been used by Alonso and Camara (2016) to study Bayesian persuasion with heterogenous priors; see also Board and Lu (2018).

The choice of the regularization scheme. We can envision multiple ways to "regularize" a cost function c and make it experimental. We believe, however, that the particular scheme $c \mapsto c^*$ we propose has several advantages, especially for applications. First, it is often easy to compute c^* starting from knowledge of c (see Examples 4 and 5). Second, c^* inherits some of the main properties of c—for example, if c is canonical, then c^* is canonical; if c is posterior separable, then c^* is posterior separable (see Section 5 in the online appendix). Third, the regularization scheme $c \mapsto c^*$ works particularly well for the widespread entropic specification c_R. As shown by Matějka and McKay (2015), when the cost of information is c_R, the behavior of the decision maker follows a logit rule. The logit rule is one of the main tools in applications of rational inattention. As we show in Lemma 5, when the cost of information is c_R^*, the behavior of the decision maker also follows a logit rule.

The choice of the topology for experiments. In choosing a topology for experiments, we identified $\mathcal{E}$ with the function space $\Delta(X)^\Theta$. The choice is motivated by the important case in which signals provide action recommendations. If A is a finite set of actions and $X = A$, then it is convenient to identify $\mathcal{E}$ with $\Delta(A)^\Theta$, a convex subset of a Euclidean space. With this identification, the primal information acquisition problem can be solved as a finite-dimensional convex program (see Example 9).

The literature has considered other topologies for experiments, which often violate the identification between $\mathcal{E}$ and $\Delta(X)^\Theta$. For example, in statistical decision theory it is common to adopt the following notion of convergence: a sequence of experiments (P_n) converges to P if the sequence of meta-posteriors $(B(\pi^*, P_n))$ converges to $B(\pi^*, P)$. This notion of convergence does not distinguish between experiments that are Blackwell equivalent. Thus it allows to identify $\mathcal{E}$ only with a quotient space of $\Delta(X)^\Theta$.

Appendix

A Properties of the Bayes map

We review some properties of the Bayes map $B : \Delta \times \mathcal{E} \to \Delta^2$ that we use throughout the appendix. Most results are known in statistical decision theory; they can be found scattered in monographs such as Torgersen (1991). For the reader's convenience, here we provide a self-contained presentation. We present results first; proofs follows in Section A.1.

We start by discussing the algebraic properties of B. Let $B_P : \Delta \to \Delta^2$ and $B_\pi : \mathcal{E} \to \Delta^2$ be its P-section and π-section, respectively.

Lemma 7. *(i) The range of B_π is Δ^2_π. (ii) The function B_P is injective.*

Property (i) can be decomposed into two parts. First, (i) states that the range of B_π is included by Δ^2_π. This is an expression of the so called "martingale property" of Bayesian updating: the expected posterior is equal to the prior, i.e., the barycenter of $B(\pi, P)$ is π itself. Second, (i) states that Δ^2_π is included by the range of B_π. It comes from the richness assumption on the set of experiments. Properties (i) and (ii) imply that the map $\mu \mapsto \bar{\mu}$ is the (left) inverse of B_P.

By Lemma 7 every random posterior is generated by some experiment. In particular,

$$\Delta^2_+ = \{B(\pi, P) : P \in \mathcal{E} \text{ and } \pi \in \Delta_+\};$$

that is, Δ^2_+ consists of all random posteriors that can be induced, via experimentation, by some prior with full support.

We turn to the ordinal properties of the Bayes map. Via this map, the Blackwell and convex orders are related in the next result, which extends to priors with full support a classic result for uniform priors.

Lemma 8. *For any $P, Q \in \mathcal{E}$, the following conditions are equivalent:*

(i) $P \succeq_b Q$;

(ii) $\int \psi\,(\mathrm{d}P/\mathrm{d}\lambda)\,\mathrm{d}\lambda \geq \int \psi\,(\mathrm{d}Q/\mathrm{d}\lambda)\,\mathrm{d}\lambda$ *for all* $\psi \in Cs(\mathbb{R}^\Theta_+)$;

(iii) $B(\pi, P) \succeq_{cv} B(\pi, Q)$ *for some* $\pi \in \Delta_+$;

(iv) $B(\pi, P) \succeq_{cv} B(\pi, Q)$ *for all* $\pi \in \Delta_+$.

An implication of this result is that, for all $\pi \in \Delta_+$ and $\mu \in \Delta^2_\pi$, the set $B^{-1}_\pi(\mu)$ is an equivalence class in the Blackwell order. When π does not have full support, we only have that (i) implies (iv). Since sublinear functions are positively homogeneous, the value of the integrals in (ii) is independent of the control measure λ (indeed, the Blackwell order depends only on the distribution of the likelihood ratios).

The proof of Lemma 8 builds on the following simple result that we will often use throughout the appendix. To state the result, let $\Theta = \{1, ..., n\}$ and $w = (w_1, ..., w_n) \in \mathbb{R}^n_+$. A self-map $\psi \longmapsto \psi_w$ on the space $Cs(\mathbb{R}^n_+)$ is defined by

$$\psi_w(z_1, ..., z_n) = \psi(w_1 z_1, ..., w_n z_n) \qquad \forall z \in \mathbb{R}^n_+.$$

Denote by $\hat{\phi} \in Cs(\mathbb{R}^n_+)$ the sublinear extension of $\phi \in Cv(\Delta)$:

$$\hat{\phi}(z) = \begin{cases} (\sum_i z_i)\phi\left(\frac{z_1}{\sum_i z_i}, \ldots, \frac{z_n}{\sum_i z_i}\right) & \text{if } \sum_i z_i > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Lemma 9. *If* $\mu = B(\pi, P)$, *then for all* $\phi \in Cv(\Delta)$

$$\int_\Delta \phi(p)\, \mathrm{d}\mu(p) = \int_X \hat{\phi}_\pi \left(\frac{\mathrm{d}P(x)}{\mathrm{d}\lambda(x)}\right) \mathrm{d}\lambda(x).$$

The result can be read as a change of variables which relates integrals over posteriors to integrals over signal realizations. The result generalizes to the case in which $\phi : \Delta \to (-\infty, \infty]$ is convex and lower semicontinuous.

We now discuss the convexity properties of the Bayes map.

Lemma 10. *For* $\alpha \in [0, 1]$, $\pi, \rho \in \Delta$, *and* $P, Q \in \mathcal{E}$, *the following conditions hold:*

(i) $\alpha B(\pi, P) + (1 - \alpha)B(\rho, P) \succeq_{cv} B(\alpha\pi + (1 - \alpha)\rho, P)$;

(ii) $\alpha B(\pi, P) + (1 - \alpha)B(\pi, Q) \succeq_{cv} B(\pi, \alpha P + (1 - \alpha)Q)$.

The next technical lemma constructs regions where the Bayes map is affine.

Lemma 11. *Let* $\pi \in \Delta$ *and* $\mu, \nu \in \Delta^2_\pi$. *There are* $P, Q \in \mathcal{E}$ *such that (i)* $B(\pi, P) = \mu$, *(ii)* $B(\pi, Q) = \nu$, *and (iii)* $B(\pi, \alpha P + (1 - \alpha)Q) = \alpha\mu + (1 - \alpha)\nu$ *for all* $\alpha \in [0, 1]$.

The proof of the result uses the richness of the signal space to construct P and Q with the desired properties.

We conclude the section by addressing the continuity properties of the Bayes map.

Lemma 12. *(i). If $\pi_n \to \pi$ and $B(\pi^*, P_n) \to B(\pi^*, P)$, then $B(\pi_n, P_n) \to B(\pi, P)$.*

(ii). If $P_n \to P$ and $B(\pi, P_n) \to \mu$, then $\mu \succeq_{cv} B(\pi, P)$.

The next technical lemma constructs regions where the section B_π is continuous.

Lemma 13. *Let (μ_n) be a sequence in Δ^2_π with $\mu_n \to \mu$. There is a sequence of experiments (P_n) such that (i) $P_n \to P$, (ii) $B(\pi, P_n) = \mu_n$ for every n, and (iii) $B(\pi, P) = \mu$.*

The proof of the result uses the richness of the signal space to construct (P_n) and P with the desired properties.

A.1 Proofs

Proof of Lemma 7. (i). If $\mu = B(\pi, P)$, then

$$\bar{\mu}(\theta) = \int_X \frac{\pi(\theta)\frac{\mathrm{d}P_\theta(x)}{\mathrm{d}\lambda(x)}}{\sum_\tau \pi(\tau)\frac{\mathrm{d}P_\tau(x)}{\mathrm{d}\lambda(x)}} \,\mathrm{d}P_\pi(x) = \pi(\theta) \int_X \frac{\mathrm{d}P_\theta(x)}{\mathrm{d}\lambda(x)} \,\mathrm{d}\lambda(x) = \pi(\theta).$$

Conversely, for $\mu \in \Delta^2_\pi$, take the standard experiment $P : \Theta \to \Delta$ given by $\mathrm{d}P/\mathrm{d}\mu = p/\bar{\mu}$. By hypothesis P belongs to $\mathcal{E}$. It is easy to verify that $B_\pi(P) = \mu$. Overall, we conclude that Δ^2_π is the range of B_π.

(ii). $B_P(\pi) = B_P(\rho)$ implies $\overline{B_P(\pi)} = \overline{B_P(\rho)}$, which implies $\pi = \rho$ by (i). ■

Proof of Lemma 9. If P is simple, then μ has finite support and

$$\begin{aligned}\sum_p \phi(p)\mu(p) &= \sum_x \phi\left(p_x\right) P_\pi\left(x\right) = \sum_x \phi\left(\frac{\pi_1 P_1(x)}{P_\pi\left(x\right)}, ..., \frac{\pi_n P_n(x)}{P_\pi\left(x\right)}\right) P_\pi\left(x\right) \\ &= \sum_x \hat{\phi}\left(\pi_1 P_1(x), ..., \pi_n P_n(x)\right) = \sum_x \hat{\phi}_\pi\left(P_1(x), ..., P_n(x)\right).\end{aligned}$$

If P is not simple, the argument is easily extended via the densities $\mathrm{d}P/\mathrm{d}\lambda$. ■

Proof of Lemma 8. We first prove that (ii) and (iii) are equivalent. If $\pi \in \Delta_+$, then $Cs(\mathbb{R}^\Theta_+) = \{\psi_\pi : \psi \in Cs(\mathbb{R}^\Theta_+)\}$. Moreover, $Cs(\mathbb{R}^\Theta_+) = \left\{\hat{\phi} : \phi \in Cv(\Delta)\right\}$. Thus $Cs(\mathbb{R}^\Theta_+) = \left\{\hat{\phi}_\pi : \phi \in Cv(\Delta)\right\}$. The equivalence of (ii) and (iii) then follows from Lemma 9. Since the choice of $\pi \in \Delta_+$ was arbitrary, we deduce that (ii)-(iv) are equivalent.

A classic result establishes the equivalence of (i) and (iii) for the uniform prior π^* (see, e.g., Blackwell and Girshick, 1954, p. 328, for the finite case and Theorem 1 of Le Cam, 1996). So, we can write

$$\begin{aligned} P \succeq_b Q &\iff B(\pi^*, P) \succeq_{cv} B(\pi^*, Q) \\ &\iff \int_X \psi\left(\frac{\mathrm{d}P}{\mathrm{d}\lambda}\right)\mathrm{d}\lambda \geq \int_X \psi\left(\frac{\mathrm{d}Q}{\mathrm{d}\lambda}\right)\mathrm{d}\lambda \qquad \forall \psi \in Cs(\mathbb{R}^n_+) \\ &\iff B(\pi, P) \succeq_{cv} B(\pi, Q) \end{aligned}$$

where the last two equivalences follow from the equivalence of (ii) and (iii), in the first case when applied to π^*. ■

Proof of Lemma 10. (i). Let $\mu = B(\alpha\pi + (1-\alpha)\rho, P)$, $\nu_1 = B(\pi, P)$ and $\nu_2 = B(\rho, P)$. For every $\psi \in Cs(\mathbb{R}^\Theta_+)$

$$\psi_{\alpha\pi+(1-\alpha)\rho}(p) \leq \alpha\psi_\pi(p) + (1-\alpha)\psi_\rho(p) \qquad \forall p \in \Delta.$$

Thus, for all $\phi \in Cv(\Delta)$, from Lemma 9 we obtain

$$\begin{aligned} \int_\Delta \phi\,\mathrm{d}\mu &= \int_X \hat{\phi}_{\alpha\pi+(1-\alpha)\rho}\left(\frac{\mathrm{d}P}{\mathrm{d}\lambda}\right)\mathrm{d}\lambda \leq \alpha \int_X \hat{\phi}_\pi\left(\frac{\mathrm{d}P}{\mathrm{d}\lambda}\right)\mathrm{d}\lambda + (1-\alpha)\int_X \hat{\phi}_\rho\left(\frac{\mathrm{d}P}{\mathrm{d}\lambda}\right)\mathrm{d}\lambda \\ &= \alpha\int_\Delta \phi\,\mathrm{d}\nu_1 + (1-\alpha)\int_\Delta \phi\,\mathrm{d}\nu_2 = \int_\Delta \phi\,\mathrm{d}(\alpha\nu_1 + (1-\alpha)\nu_2). \end{aligned}$$

We conclude that $\alpha\nu_1 + (1-\alpha)\nu_2 \succeq_{cv} \mu$.

(ii). Let $\mu = B(\pi, \alpha P + (1-\alpha)Q)$, $\nu_1 = B(\pi, P)$ and $\nu_2 = B(\pi, Q)$. For all $\phi \in Cv(\Delta)$, from Lemma 9 we obtain

$$\begin{aligned} \int_\Delta \phi\,\mathrm{d}\mu &= \int_X \hat{\phi}_\pi\left(\frac{\mathrm{d}(\alpha P + (1-\alpha)Q)}{\mathrm{d}\lambda}\right)\mathrm{d}\lambda = \int_X \hat{\phi}_\pi\left(\alpha\frac{\mathrm{d}P}{\mathrm{d}\lambda} + (1-\alpha)\frac{\mathrm{d}Q}{\mathrm{d}\lambda}\right)\mathrm{d}\lambda \\ &\leq \alpha\int_X \hat{\phi}_\pi\left(\frac{\mathrm{d}P}{\mathrm{d}\lambda}\right)\mathrm{d}\lambda + (1-\alpha)\int_X \hat{\phi}_\pi\left(\frac{\mathrm{d}Q}{\mathrm{d}\lambda}\right)\mathrm{d}\lambda \\ &= \alpha\int_\Delta \phi\,\mathrm{d}\nu_1 + (1-\alpha)\int_\Delta \phi\,\mathrm{d}\nu_2 = \int_\Delta \phi\,\mathrm{d}(\alpha\nu_1 + (1-\alpha)\nu_2). \end{aligned}$$

We conclude that $\alpha\nu_1 + (1-\alpha)\nu_2 \succeq_{cv} \mu$. ■

Proof of Lemma 11. Let $\hat{P}, \hat{Q} \in \mathcal{E}$ such that $B(\pi, \hat{P}) = \mu$ and $B(\pi, \hat{Q}) = \nu$. Let $Y = \{0, 1\}$ and $Z = X \times Y$. Define $\tilde{P}, \tilde{Q} : \Theta \to \Delta(Z)$ by $\tilde{P}_\theta(A \times \{0\}) = \hat{P}_\theta(A)$ and $\tilde{Q}_\theta(A \times \{1\}) = \hat{Q}_\theta(A)$. It is easy to check that $B(\pi, \tilde{P}) = \mu$, $B(\pi, \tilde{Q}) = \nu$, and $B(\pi, \alpha\tilde{P} + (1-\alpha)\tilde{Q}) = \alpha\mu + (1-\alpha)\nu$.[22]

Being X and Z Polish spaces, there is a Borel isomorphism $T : Z \to X$. Define $P, Q \in \mathcal{E}$

[22]Intuitively, $\alpha\tilde{P} + (1-\alpha)\tilde{Q}$ represents the experiment where a coin with bias α is tossed and, depending on where the coin lands, experiment $\hat{P}$ or $\hat{Q}$ is performed.

by $P_\theta = \tilde{P}_\theta \circ T^{-1}$ and $Q = \tilde{Q} \circ T^{-1}$. We have

$$\left(\alpha\tilde{P}_\theta + (1-\alpha)\tilde{Q}_\theta\right) \circ T^{-1} = \alpha\left(\tilde{P}_\theta \circ T^{-1}\right) + (1-\alpha)\left(\tilde{Q}_\theta \circ T^{-1}\right) = \alpha P_\theta + (1-\alpha)Q_\theta.$$

Thus $B(\pi, P) = \mu$, $B(\pi, Q) = \nu$, and $B(\pi, \alpha P + (1-\alpha)Q) = \alpha\mu + (1-\alpha)\nu$. ∎

Proof of Lemma 12. (i). For every n, take $\mu_n = B(\pi_n, P_n)$ and $\mu_n^* = B(\pi^*, P_n)$. Take also $\mu = B(\pi, P)$ and $\mu^* = B(\pi^*, P)$. Being Δ_π^2 compact, we can assume without loss of generality that $B(\pi_n, P_n) \to \nu$ for some random posterior ν. Let $\phi \in Cv(\Delta)$. From Lemma 9 for every n

$$\int_\Delta \phi(p)\,\mathrm{d}\mu_n(p) = \int_X \hat{\phi}_{\pi_n}(\mathrm{d}P_n(x)/\mathrm{d}\lambda(x))\,\mathrm{d}\lambda(x) = \int_\Delta \hat{\phi}_{\frac{\pi_n}{\pi^*}}(p)\,\mathrm{d}\mu_n^*(p).$$

The map $(\rho, p) \longmapsto \hat{\phi}_{\frac{\rho}{\pi^*}}(p)$ is jointly continuous. Thus, from $\mu_n^* \to \mu^*$, we obtain

$$\int_\Delta \phi\,\mathrm{d}\nu = \lim_n \int_\Delta \phi\,\mathrm{d}\nu_n = \lim_n \int_\Delta \hat{\phi}_{\frac{\pi_n}{\pi^*}}\,\mathrm{d}\mu_n^* = \int_\Delta \hat{\phi}_{\frac{\pi}{\pi^*}}\,\mathrm{d}\mu^* = \int_\Delta \phi\,\mathrm{d}\mu.$$

Since ϕ is arbitrary, we conclude that $\mu \sim_{cv} \mu$, i.e., $\mu = \nu$.

(ii). Set $\nu = B(\pi, P)$ and take $\phi \in Cv(\Delta)$: we wish to show that $\int_\Delta \phi\,\mathrm{d}\mu \geq \int_\Delta \phi\,\mathrm{d}\nu$. Let $\Theta = \{1, \ldots, k\}$ and $p = (p_1, \ldots, p_k) \in \Delta$. Assume that ϕ is piecewise linear (the general case follows by approximation): there exists a finite set $Z \subseteq \mathbb{R}^k$ such that

$$\phi(p) = \max_{z \in Z} \sum_i z_i p_i.$$

For every n, define $\mu_n = B(\pi, P_n)$. We have $\int_\Delta \phi\,\mathrm{d}\mu_n \to \int_\Delta \phi\,\mathrm{d}\mu$ and

$$\int_\Delta \phi(p)\,\mathrm{d}\mu_n(p) = \int_X \left(\max_{z \in Z} \sum_i z_i \frac{\mathrm{d}P_{n,i}(x)}{\mathrm{d}\lambda(x)}\right) \mathrm{d}\lambda(x).$$

Let $B(X, Z)$ be the space of bounded measurable functions from X into Z. For every $f \in B(X, Z)$

$$\int_X \left(\max_{z \in Z} \sum_i z_i \frac{\mathrm{d}P_{n,i}(x)}{\mathrm{d}\lambda(x)}\right) \mathrm{d}\lambda(x) \geq \sum_i \int_X f_i(x)\,\mathrm{d}P_{n,i}(x).$$

Conversely, by the Measurable Maximum Theorem (Aliprantis and Border, 2006, Theorem 18.19), there is a measurable function $f_n : X \to Z$ such that

$$\int_X \left(\max_{z \in Z} \sum_i z_i \frac{\mathrm{d}P_{n,i}(x)}{\mathrm{d}\lambda(x)}\right) \mathrm{d}\lambda(x) = \sum_i \int_X f_{n,i}(x)\,\mathrm{d}P_{n,i}(x).$$

Let $C(X,Z)$ be the space of continuous functions from X into Z. By a standard approximation argument, there is a sequence $(f_{n,m})$ in $C(X,Z)$ such that

$$\sum_i \int_X f_{n,i}(x)\,\mathrm{d}P_{n,i}(x) = \sup_m \sum_i \int_X f_{n,m,i}(x)\,\mathrm{d}P_{n,i}(x).$$

Overall, we obtain for all n

$$\int_\Delta \phi(p)\,\mathrm{d}\mu_n(p) = \sup_{f\in C(X,Z)} \sum_i \int_X f_i(x)\,\mathrm{d}P_{n,i}(x)$$

The analogous equality holds for μ and P. Thus

$$\begin{aligned}\int_\Delta \phi(p)\,\mathrm{d}\mu(p) &= \lim_n \sup_{f\in C(X,Z)} \sum_i \int_X f_i(x)\,\mathrm{d}P_{n,i}(x)\\ &\geq \sup_{f\in C(X,Z)} \lim_n \sum_i \int_X f_i(x)\,\mathrm{d}P_{n,i}(x)\\ &= \sup_{f\in C(X,Z)} \sum_i \int_X f_i(x)\,\mathrm{d}P_i(x) = \int_\Delta \phi(p)\,\mathrm{d}\nu(p).\end{aligned}$$

■

Proof of Lemma 13. For every n, let P_n be the standard experiment given by $\frac{\mathrm{d}P_n}{\mathrm{d}\mu_n} = \frac{p}{\pi}$. Let P be the standard experiment given by $\frac{\mathrm{d}P}{\mathrm{d}\mu} = \frac{p}{\pi}$. We have $B(\pi,P)=\mu$ and , for every n, $B(\pi,P_n)=\mu_n$. Moreover, for every continuous function $\phi:\Delta\to\mathbb{R}$ and $\theta\in\Theta$

$$\begin{aligned}\int_\Delta \phi(p)\,\mathrm{d}P_{n,\theta}(p) &= \int_\Delta \phi(p)\frac{p(\theta)}{\pi(\theta)}\,\mathrm{d}\mu_n(p)\\ &\to \int_\Delta \phi(p)\frac{p(\theta)}{\pi(\theta)}\,\mathrm{d}\mu(p) = \int_\Delta \varphi(p)\,\mathrm{d}P_\theta(p).\end{aligned}$$

Thus $P_n \to P$. ■

B Proofs of the results in the main text

Proof of Lemma 1. Reflexivity and transitivity are inherited from the Blackwell order.

(i). "If." Let $P^\mu, P^\nu \in \mathcal{E}$ such that $B(\bar\mu, P^\mu)=\mu$ and $\nu = B(\bar\nu, P^\nu)$. By hypothesis,

$$\{P\in\mathcal{E} : B(\bar\mu,P)\succeq_{cv}\mu\} \subseteq \{Q\in\mathcal{E} : B(\bar\nu,Q)\succeq_{cv}\nu\}.$$

Thus $B(\bar\nu,P^\mu)\succeq_{cv} B(\bar\nu,P^\nu)$. By Lemma 8 there is a stochastic kernel K such that $P^\nu_\theta = KP^\mu_\theta$ for all $\theta\in\operatorname{supp}\bar\nu$. Define $Q\in\mathcal{E}$ by $Q_\theta = KP^\mu_\theta$ for all θ. We have $P^\mu \succeq_b Q$ and

$B(\bar{\nu}, Q) = \nu$ (being $Q_\theta = P^\nu_\theta$ for all $\theta \in \operatorname{supp} \bar{\nu}$). Since the choice of P^μ if arbitrary, we conclude that $\mu \succeq_{ex} \nu$.

"Only if." Let $\mu, \nu \in \Delta^2$ such that $\mu \succeq_{ex} \nu$. Take $P, P^\mu \in \mathcal{E}$ be such that $B(\bar{\mu}, P) \succeq_{cv} \mu$ and $\mu = B(\bar{\mu}, P^\mu)$. By Lemma 8 there is a stochastic kernel K such that $P^\mu_\theta = KP_\theta$ for all $\theta \in \operatorname{supp} \bar{\mu}$. Define $Q' \in \mathcal{E}$ by $Q'_\theta = KP_\theta$ for all θ. We have $P \succeq_b Q'$ and $B(\bar{\mu}, Q') = \mu$. Because $\mu \succeq_{ex} \nu$, we can find $Q \in \mathcal{E}$ such that $Q' \succeq_b Q$ and $B(\bar{\nu}, Q) = \nu$. Thus $P \succeq_b Q$, which implies $B(\bar{\nu}, P) \succeq_{cv} \nu$ by Lemma 8. Since the choice of P is arbitrary, we conclude that

$$\{P \in \mathcal{E} : B(\bar{\mu}, P) \succeq_{cv} \mu\} \subseteq \{Q \in \mathcal{E} : B(\bar{\nu}, Q) \succeq_{cv} \nu\}.$$

(ii). The proof of the "if" statement is straightforward. To prove the "only if" statement, suppose that $\mu \sim_{ex} \nu$. By (i),

$$\{P \in \mathcal{E} : B(\bar{\mu}, P) \succeq_{cv} \mu\} = \{P \in \mathcal{E} : B(\bar{\nu}, P) \succeq_{cv} \nu\}. \tag{14}$$

We claim that $\bar{\mu}$ and $\bar{\nu}$ have the same support or $\mu = \delta_{\bar{\mu}}$ and $\nu = \delta_{\bar{\nu}}$. By contraposition, suppose that $\mu \neq \delta_{\bar{\mu}}$ and there is θ such that $\bar{\mu}(\theta) > 0 = \bar{\nu}(\theta)$. Take P such that $B(\bar{\mu}, P) = \mu$. If $B(\bar{\nu}, P) \not\succeq_{cv} \nu$, then (14) is false and we are done. Assume therefore that $B(\bar{\nu}, P) \succeq_{cv} \nu$. Since $\mu \neq \delta_{\bar{\mu}}$, there is $\tau \in \Theta$ such $\bar{\mu}(\tau) > 0$ and $P_\theta \neq P_\tau$. Let Q be equal to P except for $Q_\theta = P_\tau$. Since $\bar{\nu}(\theta) = 0$, $B(\bar{\nu}, Q) \succeq_{cv} \nu$. Let $\pi \in \Delta$ such that $\pi(\theta) = \pi(\tau) = 1/2$. Since $P_\theta \neq P_\tau$ and $Q_\theta = Q_\tau$, $B(\pi, P) \succ_{cv} \delta_\pi = B(\pi, Q)$. By Lemma 8, $B(\bar{\mu}, Q) \not\succeq_{cv} B(\bar{\mu}, P) = \mu$. This negates (14): we are done.

Now consider the case in which $\bar{\mu}$ and $\bar{\nu}$ have the same support. Let P^μ and P^ν be experiments that generate μ and ν. By (14), $B(\bar{\mu}, P^\nu) \succeq_{cv} B(\bar{\mu}, P^\mu)$ and $B(\bar{\nu}, P^\mu) \succeq_{cv} B(\bar{\nu}, P^\nu)$. Since $\bar{\mu}$ and $\bar{\nu}$ have the same support, it follows from Lemma 8 that $B(\bar{\mu}, P^\nu) = B(\bar{\mu}, P^\mu)$ and $B(\bar{\nu}, P^\mu) = B(\bar{\nu}, P^\nu)$. This shows that

$$\{P \in \mathcal{E} : B(\bar{\mu}, P) = \mu\} = \{P \in \mathcal{E} : B(\bar{\nu}, P) = \nu\}.$$

If instead $\mu = \delta_{\bar{\mu}}$ and $\nu = \delta_{\bar{\nu}}$, then trivially

$$\{P \in \mathcal{E} : B(\bar{\mu}, P) = \mu\} = \{P \in \mathcal{E} : P \text{ is uninformative}\} = \{P \in \mathcal{E} : B(\bar{\nu}, P) = \nu\}.$$

This concludes the proof of the "only if" direction.

(iii). Assume that $\mu \succeq_{cv} \nu$. Take $P \in \mathcal{E}$ such that $B(\bar{\mu}, P) \succeq_{cv} \mu$. Being $\mu \succeq_{cv} \nu$, we have $\bar{\mu} = \bar{\nu}$ and $B(\bar{\mu}, P) = B(\bar{\nu}, P) \succeq_{cv} \nu$. It follows from (i) that $\mu \succeq_{ex} \nu$. To prove the converse, take $P^\mu \in \mathcal{E}$ such that $B(\bar{\mu}, P^\mu) = \mu$. Because $\mu \succeq_{ex} \nu$ we have $B(\bar{\nu}, P^\mu) \succeq_{cv} \nu$ by (i). Since $\bar{\mu} = \bar{\nu}$, we conclude that $\mu = B(\bar{\nu}, P^\mu) \succeq_{cv} \nu$. ■

Proof of Lemma 2. (i) implies (ii). Let $P \succeq_b Q$. We want to show that, given any $\pi, \rho \in \Delta_+$, we have $B(\pi, P) \succeq_{ex} B(\rho, Q)$. In view of Lemma 1-(i), $B(\pi, P) \succeq_{ex} B(\rho, Q)$ if and only if

$$B(\pi, P') \succeq_{cv} B(\pi, P) \Longrightarrow B(\rho, P') \succeq_{cv} B(\rho, Q) \qquad \forall P' \in \mathcal{E}. \tag{15}$$

Let $P' \in \mathcal{E}$ be such that $B(\pi, P') \succeq_{cv} B(\pi, P)$. Since π has full support, $P' \succeq_b P$ by Lemma 8. This in turn implies $P' \succeq_b Q$ (being $P \succeq_b Q$). Thus $B(\rho, P') \succeq_{cv} B(\rho, Q)$ as desired.

(iii) implies (i). Let $P, Q \in \mathcal{E}$. Assume $B(\pi, P) \succeq_{ex} B(\rho, Q)$ for some $\pi, \rho \in \Delta_+$. Since $\succeq_{cv}$ is reflexive, we have $B(\pi, P) \succeq_{cv} B(\pi, P)$ and so, by (15), $B(\rho, P) \succeq_{cv} B(\rho, Q)$. By Lemma 8, this implies $P \succeq_b Q$, having ρ full support. Since (ii) trivially implies (iii), this completes the proof. ■

Proof of Lemma 3. Let P and Q be the standard experiments given by $\frac{\mathrm{d}P}{\mathrm{d}\mu} = \frac{p}{\bar{\mu}}$ and $\frac{\mathrm{d}Q}{\mathrm{d}\nu} = \frac{p}{\bar{\nu}}$. The experiments P and Q generate μ and ν, respectively. By Lemma 2, we have $\mu \succeq_{ex} \nu$ if and only $P \succeq_b Q$. The desired result follows from the variational representation of the Blackwell order, Lemma 8-(ii). ■

Proof of Theorem 1. (i). If c is experimental, then it is invariant under $\sim_{ex}$ by Lemma 1-(ii). Conversely, suppose that c is invariant under $\sim_{ex}$. Since $\succeq_b$ is reflexive, it follows from Lemma 2 that, for all $P \in \mathcal{E}$ and $\pi, \rho \in \Delta_+$, $B(\pi, P) \sim_{ex} B(\rho, P)$. Since c is invariant under $\sim_{ex}$, we then have

$$c\left(B(\pi, P)\right) = c\left(B(\rho, P)\right). \tag{16}$$

Fix $\pi \in \Delta_+$ and define $h_\pi : \mathcal{E} \to [0, \infty]$ by $h_\pi(P) = c(B(\pi, P))$. By Lemma 8, h_π is invariant under $\sim_b$. Along with (16) and the invariance of c under $\succeq_{ex}$, we obtain for each $\mu \in \Delta^2_+$

$$\begin{aligned} h_\pi(P^\mu) &\geq \inf\{h_\pi(P) : B(\bar{\mu}, P) = \mu\} = \inf\{c(B(\pi, P)) : B(\bar{\mu}, P) = \mu\} \\ &= \inf\{c(B(\bar{\mu}, P)) : B(\bar{\mu}, P) = \mu\} = c\left(\mu\right) = c\left(B(\bar{\mu}, P^\mu)\right) \\ &= c\left(B(\pi, P^\mu)\right) = h_\pi(P^\mu) \end{aligned}$$

where P^μ satisfies $B(\bar{\mu}, P^\mu) = \mu$. We conclude that

$$\inf\{h_\pi(P) : B(\bar{\mu}, P) = \mu\} = c\left(\mu\right) \qquad \forall \mu \in \Delta^2_+.$$

This shows that c is experimental and it is rationalized by h_π. Note that π was arbitrarily chosen in Δ_+. In particular, we can choose the uniform prior π^*.

Finally, if $h' : \mathcal{E} \to [0, \infty]$ is invariant under $\sim_b$ and induces c, then by Lemma 8

$$h_\pi(P) = c(B(\pi, P)) = \inf\{h'(Q) : B(\pi, Q) = B(\pi, P)\} = h'(P).$$

Thus $h' = h_\pi$ as desired.

(ii). Suppose that h is Blackwell monotone. Trivially

$$c_h(\mu) \geq \inf\{h(P) : B(\bar{\mu}, P) \succeq_{cv} \mu\}.$$

Now take $P, P^\mu \in \mathcal{E}$ such that $B(\bar{\mu}, P) \succeq_{cv} \mu = B(\bar{\mu}, P^\mu)$. By Lemma 8, $P \succeq_b P^\mu$. Being h Blackwell monotone, $h(P) \geq h(P^\mu)$. Since the choice of P is arbitrary, we obtain

$$c_h(\mu) \leq \inf\{h(P) : B(\bar{\mu}, P) \succeq_{cv} \mu\}.$$

We conclude that

$$c_h(\mu) = \inf\{h(P) : B(\bar{\mu}, P) \succeq_{cv} \mu\}.$$

It follows from Lemma 1-(i) that c_h is monotone in the experimental order.

(iii). Suppose that c is monotone in the convex order and experimental. By (i), $c = c_{h_c}$. Since c is monotone in the convex order, h_c is Blackwell monotone. By (ii), c is monotone in the experimental order. Conversely, if c is monotone in the experimental order, then c is monotone in the convex order by Lemma 1-(iii) and experimental by (i). ■

Proof of Corollary 1. Let c be experimental and suppose that (i) holds. Take a random posterior $\mu \in \Delta^2_+$: we want to show that $c(\mu) = 0$. Let P^μ be an experiment that generates μ—that is, such that $\mu = B(\bar{\mu}, P^\mu)$. By Lemma 2 we have $B(\bar{\mu}, P^\mu) \sim_{ex} B(\pi, P^\mu)$ for all $\pi \in \Delta_+$. Being c invariant under $\sim_{ex}$ (Theorem 1), we obtain

$$c(\mu) = c(B(\bar{\mu}, P^\mu)) = c(B(\pi, P^\mu)) \qquad \forall \pi \in \Delta_+.$$

Thus, it follows from (i) that

$$c(\mu) = \lim_{\pi(\theta) \to 1} c(B(\pi, P^\mu)) = 0$$

as desired. This shows that (i) implies (ii). The other implication is trivial. ■

Proof of Corollary 2. Define $E : \mathcal{C}^{ex} \to \mathcal{H}^b$ by $E(c) = h_c$. The functions D and E are well defined by Theorem 1-(i). For every $P \in \mathcal{E}$, we have

$$h_{c_h}(P) = c_h(B(\pi^*, P)) = h(P)$$

where the last equality follows from h being invariant under $\sim_b$. In addition, for every $\mu \in \Delta^2_+$ we have

$$c_{h_c}(\mu) = h_c(P^\mu) = c(B(\pi^*, P^\mu)) = c(\mu).$$

where first and last equality follow from h and c being invariant under $\sim_b$ and $\sim_{ex}$. We conclude that D is bijective and $E = D^{-1}$. The cases where h and c are monotone in $\succeq_b$ and $\succeq_{ex}$ follow from Theorem 1-(ii) and Theorem 1-(iii). ■

Proof of Lemma 4. Fix $\mu \in \Delta^2_+$. Let $P^\mu \in \mathcal{E}$ such that $B(\bar{\mu}, P^\mu) = \mu$. Set $\mu^* = B(\pi^*, P^\mu)$. Since the Blackwell order is reflexive, $P^\mu \sim_b P^\mu$. By Lemma 2, $\mu^* = B(\pi^*, P^\mu) \sim_{ex} B(\bar{\mu}, P^\mu) = \mu$. Such μ^* is unique. Indeed, let $\nu^* \in \Delta^2_{\pi^*}$ such that $\mu \sim_{ex} \nu^*$. By the transitivity of $\sim_{ex}$, we have $\mu^* \sim_{ex} \nu^*$. Since they have the same barycenter π^*, by Lemma 1-(ii) we have $\mu^* \sim_{cv} \nu^*$, so that $\mu^* = \nu^*$. We conclude that the map $\Delta^2_+ \ni \mu \mapsto \mu^* \in \Delta^2_{\pi^*}$ is well defined.

(i). Let $\mu, \nu \in \Delta^2_+$. By the transitivity of $\succeq_{ex}$, we have $\nu \succeq_{ex} \mu$ if and only if $\nu^* \succeq_{ex} \mu^*$. By Lemma 1-(ii), we have $\nu^* \succeq_{ex} \mu^*$ if and only if $\nu^* \succeq_{cv} \mu^*$.

(ii). This follows from c being invariant under $\sim_{ex}$ (Theorem 1). ■

Proof of Corollary 3. Let $\mu, \nu \in \Delta^2_+$. If $\nu \sim_{ex} \mu$, then, by Lemma 4-(i), $\nu^* = \mu^*$ and so

$$c^*(\nu) = c(\nu^*) = c(\mu^*) = c^*(\mu).$$

By Theorem 1 the cost function c^* is experimental. Assume, in addition, that c is monotone in the convex order. If $\nu \succeq_{cv} \mu$, then $\nu^* \succeq_{cv} \mu^*$ by Lemma 4-(i). Thus

$$c^*(\nu) = c(\nu^*) \geq c(\mu^*) = c^*(\mu),$$

being c monotone in the convex order. ■

Proof of Lemma 5. The buyer's problem of best responding to σ can be rewritten as

$$\max_{P,\beta} \sum_{t,x} \left(\frac{(v-t)\sigma(t)}{\sigma^*(t)} \right) \beta(x) P_t(x) \sigma^*(t) - k \sum_t D_{KL}(P_t \| P_{\sigma^*}) \sigma^*(t).$$

That is formally equivalent to a rational inattention problem where the utility from purchasing the good is $\frac{(v-t)\sigma(t)}{\sigma^*(t)}$, the buyer's prior is σ^*, and the cost is $c(B(\sigma^*, P)) = kc_R(B(\sigma^*, P))$. Thus we can apply the results of Matějka and McKay (2015) and obtain the desired logit representation. ■

Proof of Proposition 4. Let (σ, P, β) an equilibrium with $\beta_\sigma > 0$. By Proposition 3, $\sigma(2v) > 0$ and $\sigma(\frac{2v}{3}) > 0$. The seller's indifference condition between $t = \frac{2v}{3}$ and $t = 2v$ is

$$\frac{2v}{3} \beta_{\frac{2v}{3}} = 2v \beta_{2v} \qquad \Longleftrightarrow \qquad \beta_{\frac{2v}{3}} = 3\beta_{2v}.$$

From Proposition 5, $\beta_v = \beta_{\sigma^*}$. Thus

$$\frac{1}{3}\beta_{\frac{2v}{3}} + \frac{1}{3}\beta_v + \frac{1}{3}\beta_{2v} = \beta_{\sigma^*} \qquad \Longleftrightarrow \qquad \frac{1}{2}\beta_{\frac{2v}{3}} + \frac{1}{2}\beta_{2v} = \beta_{\sigma^*}$$

It follows from $\beta_{\frac{2v}{3}} = 3\beta_{2v}$ that

$$\beta_{\frac{2v}{3}} = \frac{3}{2}\beta_{\sigma^*} \qquad \text{and} \qquad \beta_{2v} = \frac{1}{2}\beta_{\sigma^*}.$$

In particular, the seller is indifferent between all offers. From Proposition 5,

$$\frac{e^{\frac{\left(v-\frac{2v}{3}\right)\sigma\left(\frac{3v}{3}\right)}{k\sigma^*\left(\frac{2v}{3}\right)}}\beta_{\sigma^*}}{e^{\frac{\left(v-\frac{2v}{3}\right)\sigma\left(\frac{2v}{3}\right)}{k\sigma^*\left(\frac{2v}{3}\right)}}\beta_{\sigma^*} + 1 - \beta_{\sigma^*}} = \frac{3}{2}\beta_{\sigma^*} \qquad \text{and} \qquad \frac{e^{\frac{(v-2v)\sigma(2v)}{k\sigma^*(2v)}}\beta_{\sigma^*}}{e^{\frac{(v-2v)\sigma(2v)}{k\sigma^*(2v)}}\beta_{\sigma^*} + 1 - \beta_{\sigma^*}} = \frac{1}{2}\beta_{\sigma^*}.$$

Simple algebra shows that

$$e^{\frac{\left(v-\frac{2v}{3}\right)\sigma\left(\frac{3v}{3}\right)}{k\sigma^*\left(\frac{2v}{3}\right)}}\left(1 - \frac{3}{2}\beta_{\sigma^*}\right) = \frac{3}{2}\left(1 - \beta_{\sigma^*}\right) \qquad \text{and} \qquad e^{\frac{(v-2v)\sigma(2v)}{k\sigma^*(2v)}}\left(1 - \frac{1}{2}\beta_{\sigma^*}\right) = \frac{1}{2}(1 - \beta_{\sigma^*}).$$

Thus $\beta_{\sigma^*} \in (0, \frac{2}{3})$. Simplifying further, we obtain

$$\sigma\left(\frac{3v}{3}\right) = \frac{k}{v}\log\frac{3(1-\beta_{\sigma^*})}{2 - 3\beta_{\sigma^*}} \qquad \text{and} \qquad \sigma(2v) = \frac{k}{3v}\log\frac{2 - \beta_{\sigma^*}}{1 - \beta_{\sigma^*}}.$$

This concludes the proof of the first part of the proposition.

Now take $z \in (0, \frac{2}{3})$ such that

$$\frac{k}{v}\log\frac{3(1-z)}{2-3z} + \frac{k}{3v}\log\frac{2-z}{1-z} \leq 1. \tag{17}$$

Define $\sigma \in \Delta(T)$ by

$$\sigma(t) = \begin{cases} \frac{k}{3(v-t)}\log\frac{v(1-z)}{t-vz} & \text{if } t \neq v \\ 1 - \frac{k}{v}\log\frac{3(1-z)}{2-3z} - \frac{k}{3v}\log\frac{2-z}{1-z} & \text{if } t = v. \end{cases}$$

By (17), σ is well defined. Simple algebra shows that

$$\sum_t \sigma^*(t)\frac{e^{\frac{(v-t)\sigma(t)}{k\sigma^*(t)}}}{e^{\frac{(v-t)\sigma(t)}{k\sigma^*(t)}}z + 1 - z} = 1.$$

That is the first order condition of the maximization problem in Proposition 5. Thus we can find a best response to (P, β) to σ such that $\beta_{\sigma^*} = z$. In addition, for all $t \in T$,

$$\beta_t = \frac{e^{\frac{(v-t)\sigma(t)}{k\sigma^*(t)}} z}{e^{\frac{(v-t)\sigma(t)}{k\sigma^*(t)}} z + 1 - z} = \frac{v}{t} z.$$

Therefore the seller is indifferent between all offers: σ is a best response to (P, β) as well. We conclude that (σ, P, β) is an equilibrium with $\beta_\sigma > 0$ and $\beta_{\sigma^*} = z$.

Finally, consider the function $f : (0, \frac{2}{3}) \to \mathbb{R}_+$ given by

$$f(z) = \frac{k}{v} \log \frac{3(1-z)}{2-3z} + \frac{k}{3v} \log \frac{2-z}{1-z}.$$

Note that f is continuous and strictly increasing. In addition,

$$\lim_{z \to 0} f(z) = \frac{k}{v} \log \frac{3}{2} + \frac{k}{3v} \log 2 \qquad \text{and} \qquad \lim_{z \to \frac{2}{3}} f(z) = \infty.$$

As shown above, if (σ, P, β) is an equilibrium with $\beta_\sigma > 0$, then

$$\lim_{z \to 0} f(z) < f(\beta_{\sigma^*}) = \sigma\left(\frac{2}{3} v\right) + \sigma(v) \leq 1.$$

Conversely, if $\lim_{z \to 0} f(z) < 1$, then by Brouwer's fixed point theorem there exits z such that $f(z) \leq 1$. As shown above, this implies that there exists an equilibrium (σ, P, β) with $\beta_\sigma > 0$. The condition $\lim_{z \to 0} f(z) < 1$ is equivalent to $k < \frac{3v}{3 \log 3 - 2 \log 2}$. ■

Proof of Proposition 5. (i). Let $\alpha \in [0, 1]$ and $\mu, \nu \in \Delta^2_\pi$. By Lemma 11 we can choose P^μ and P^ν generating μ and ν, respectively, such that $B(\pi, \alpha P^\mu + (1-\alpha) P^\nu) = \alpha\mu + (1-\alpha)\nu$. Being h Blackwell monotone, $c_h(\mu) = h(P^\mu)$, $c_h(\nu) = h(P^\nu)$, and $c_h(\alpha\mu + (1-\alpha)\nu) = h(\alpha P^\mu + (1-\alpha) P^\nu)$. It follows from the convexity of h that

$$c_h(\alpha\mu + (1-\alpha)\nu) = h(\alpha P^\mu + (1-\alpha) P^\nu) \leq \alpha h(P^\mu) + (1-\alpha) h(P^\nu) = \alpha c_h(\mu) + (1-\alpha) c_h(\nu).$$

We conclude that c_h is convex.

(ii). Let $\alpha \in [0, 1]$ and $P, Q \in \mathcal{E}$. Because c is monotone in $\succeq_{ex}$, by Lemma 10

$$c(B(\pi^*, \alpha P + (1-\alpha) Q)) \leq c(\alpha B(\pi^*, P) + (1-\alpha) B(\pi^*, Q)).$$

In addition, because c is convex on $\Delta^2_{\pi^*}$,

$$c(\alpha B(\pi^*, P) + (1-\alpha) B(\pi^*, Q)) \leq \alpha c(B(\pi^*, P)) + (1-\alpha) c(B(\pi^*, Q)).$$

We conclude that

$$h_c(\alpha P + (1-\alpha)Q) = c(B(\pi^*, \alpha P + (1-\alpha)Q))$$
$$\leq \alpha c(B(\pi^*, P)) + (1-\alpha)c(B(\pi^*, Q)) = \alpha h_c(P) + (1-\alpha)h_c(Q).$$

We conclude that h_c is convex.

(iii). Since c is convex on $\Delta^2_{\pi^*}$ and c^* agrees with c on $\Delta^2_{\pi^*}$, we have that c^* is convex on $\Delta^2_{\pi^*}$. By (ii) this implies that h_{c^*} is convex. Because $c_{h_{c^*}} = c^*$ (Corollary 2), it follows from (i) that c^* is convex on each Δ^2_{π}. ■

Proof of Proposition 6. (i). Let (μ_n) be a sequence in Δ_+ with limit μ. For every n, we take ν_n such that $\nu_n \sim_{ex} \mu_n$ and $\bar{\nu}_n = \bar{\mu}$. Being $\Delta^2_{\bar{\mu}}$ compact, we can assume without loss of generality that the sequence (ν_n) converges to some $\nu \in \Delta^2_{\bar{\mu}}$. Note that $\bar{\mu}_n \to \bar{\mu}$. Thus $\nu = \mu$ by Lemma 12.

By Lemma 13, we can choose a sequence (P_n) in $\mathcal{E}$ with limit P such that $B(\bar{\mu}, P) = \nu$ and, for every n, $B(\bar{\mu}, P_n) = \nu_n$. Being h Blackwell monotone, $c_h(\nu) = h(P)$ and, for every n, $c_h(\nu_n) = h(P_n)$. Since $\mu = \nu$, we have $c_h(\mu) = c_h(\nu)$. Since c_h is monotone in the experimental order, for every n we have $c_h(\mu_n) = c_h(\nu_n)$. Then it follows from the lower semicontinuity h that

$$\liminf_n c_h(\mu_n) = \liminf_n c_h(\nu_n) = \liminf_n h(P_n) \geq h(P) = c_h(\nu) = c_h(\mu).$$

We conclude that c_h is lower semicontinuous on Δ^2_+.

(ii). Let (P_n) be a sequence in $\mathcal{E}$ with limit P. Because $\Delta^2_{\pi^*}$ is compact, we can assume without loss of generality that $B(\pi^*, P_n)$ converges to some μ. Because c is lower semicontinuous on $\Delta^2_{\pi^*}$,

$$\liminf_n h_c(P_n) = \liminf_n c(B(\pi^*, P_n)) \geq c(\mu).$$

From Lemma 12 we have $\mu \succeq_{cv} B(\pi^*, P)$. Thus, being c monotone in $\succeq_{ex}$, we have

$$c(\mu) \geq c(B(\pi^*, P)) = h_c(P).$$

We conclude that h_c is lower semicontinuous.

(iii). Since c is lower semicontinuous on $\Delta^2_{\pi^*}$ and c^* agrees with c on $\Delta^2_{\pi^*}$, we have c^* lower semicontinuous on $\Delta^2_{\pi^*}$. By (ii) we obtain that h_{c^*} is lower semicontinuous. Being $c_{h_{c^*}} = c^*$ (Corollary 2), by (i) the function c^* is lower semicontinuous on Δ^2_+. ■

Proof of Lemma 6. The "if" statement is trivial. To prove the "only if" statement, for every $\phi \in Cv(\Delta)$ and $\pi \in \Delta_+$ we define

$$V(\phi, \pi) = \max_{\mu \in \Delta^2_\pi} \int_\Delta \phi(p) \mathrm{d}\mu(p) - c(\mu).$$

By Theorem 2 of De Oliveira, Denti, Mihm, and Ozbek (2017), we have for every $\pi \in \Delta_+$ and $\mu \in \Delta^2_\pi$

$$c(\mu) = \sup_{\phi \in Cv(\Delta)} \int_\Delta \phi(p) \mathrm{d}\mu(p) - V(\phi, \pi).$$

In particular,

$$0 = c(\delta_\pi) = \sup_{\phi \in Cv(\Delta)} \phi(\pi) - V(\phi, \pi)$$

We obtain the desired result by setting $\Phi^\pi = \{\phi - V(\phi, \pi) : \phi \in Cv(\Delta)\}$. ■

Proof of Proposition 7. (i). "If." Assume that

$$c(\mu) = \sup_{\psi \in \Psi} \int_\Delta \psi\left(\frac{p}{\bar{\mu}}\right) \mathrm{d}\mu(p) - \sup_{\psi \in \Psi} \psi(1) \qquad \forall \mu \in \Delta^2_+.$$

On each Δ^+_π, the function c is the supremum of affine continuous functions $p \longmapsto \psi(p/\pi)$ where $\psi \in \Psi$. Thus c is convex and lower semicontinuous on each Δ^+_π. In addition,

$$c(\delta_\pi) = \sup_{\psi \in \Psi} \psi\left(\frac{\pi}{\pi}\right) - \sup_{\psi \in \Psi} \psi(1) = 0.$$

Thus c is canonical.

"Only if." Assume that c is canonical. Take Φ^{π^*} as in Lemma 6. From Corollary 3

$$c(\mu) = c^*(\mu) = c(\mu^*) = \sup_{\phi \in \Phi^{\pi^*}} \int_\Delta \phi(p)\, \mathrm{d}\mu^*(p) - \sup_{\phi \in \Phi^{\pi^*}} \phi(\pi^*) \qquad \forall \mu \in \Delta^2_+.$$

From Lemma 3, we have that for all $\phi \in \Phi^{\pi^*}$ and $\mu \in \Delta^2_+$

$$\int_\Delta \phi(p)\, \mathrm{d}\mu^*(p) = \int_\Delta \hat{\phi}_{\pi^*}\left(\frac{p}{\pi^*}\right) \mathrm{d}\mu^*(p) = \int_\Delta \hat{\phi}_{\pi^*}\left(\frac{p}{\bar{\mu}}\right) \mathrm{d}\mu(p).$$

The desired result follows by setting $\Psi = \left\{\hat{\phi}_{\pi^*} : \phi \in \Phi^{\pi^*}\right\}$.

(ii). "If." Assume that

$$h(P) = \sup_{\psi \in \Psi} \int_{\mathbb{R}^\Theta_+} \psi\left(\frac{\mathrm{d}P}{\mathrm{d}\lambda}\right) \mathrm{d}\lambda - \sup_{\psi \in \Psi} \psi(1) \qquad \forall P \in \mathcal{E}.$$

For every $\mu \in \Delta^2_+$, let P^μ be the standard experiment given by $\mathrm{d}P/\mathrm{d}\mu = p/\bar{\mu}$. Then

$$c_h(\mu) = h(P^\mu) = \sup_{\psi \in \Psi} \int_\Delta \psi\left(\frac{p}{\bar{\mu}}\right) \mathrm{d}\mu(p) - \sup_{\psi \in \Psi} \psi(1).$$

By (i), the cost function c_h is canonical. By Propositions 5 and 6, $h_{c_h} = h$ is canonical.

"Only if." Assume that h is canonical. By Propositions 5 and 6, c_h is canonical: take $\Psi \subseteq Cs(\mathbb{R}^\Theta_+)$ as in (i). Fix $P \in \mathcal{E}$ and choose $\mu^* = B(\pi^*, P)$. Because $h_{c_h} = h$, we obtain from Lemma 9 that

$$h(P) = c_h(\mu) = \sup_{\psi \in \Psi} \int_\Delta \psi\left(\frac{p}{\pi^*}\right) \mathrm{d}\mu(p) - \sup_{\psi \in \Psi} \psi(1) = \sup_{\psi \in \Psi} \int_\Delta \psi\left(\frac{\mathrm{d}P}{\mathrm{d}\lambda}\right) \mathrm{d}\lambda - \sup_{\psi \in \Psi} \psi(1).$$

The desired result follows. ■

Proof of Proposition 8. The "if" direction follows by setting $\phi^\pi(p) = \psi\left(\frac{p}{\pi}\right)$. Conversely, suppose that $c \in \mathcal{C}^{mx}$ is posterior separable. Take ϕ^{π^*} as in Definition (8). To shorten notation, we drop the superscript and write ϕ instead of ϕ^{π^*}. From Corollary 3, for all $\mu \in \Delta^2_+$,

$$c(\mu) = c^*(\mu) = c(\mu^*) = \int_\Delta \phi(p)\, \mathrm{d}\mu^*(p) - \phi(\pi^*).$$

From Lemma 3, we have that for all $\mu \in \Delta^2_+$

$$\int_\Delta \phi(p)\, \mathrm{d}\mu^*(p) = \int_\Delta \hat{\phi}_{\pi^*}\left(\frac{p}{\pi^*}\right) \mathrm{d}\mu^*(p) = \int_\Delta \hat{\phi}_{\pi^*}\left(\frac{p}{\bar{\mu}}\right) \mathrm{d}\mu(p).$$

The desired follows by setting $\psi = \hat{\phi}_{\pi^*}$. ■

Proof of Proposition 9. (i). For $\mu \in \Delta^2_+$ let P^μ be the standard experiment given by $\frac{\mathrm{d}P}{\mathrm{d}\mu} = \frac{p}{\bar{\mu}}$. Being h Blackwell monotone, we have

$$c_h(\mu) = h(P^\mu) = \int_\Delta \psi\left(\frac{p}{\bar{\mu}}\right) \mathrm{d}\mu(p) - \psi(1).$$

From the "if" direction of Proposition 8 it follows that c_h is posterior separable.

(ii). Assume c is posterior separable and take ψ as in Proposition 8. Take $P \in \mathcal{E}$ and set $\mu = B(\pi^*, P)$. From Lemma 9 we obtain

$$h_c(P) = c(\mu) = \int_\Delta \psi\left(\frac{p}{\pi^*}\right) \mathrm{d}\mu(p) - \psi(1) = \sup_{\psi \in \Psi} \int_\Delta \psi\left(\frac{\mathrm{d}P}{\mathrm{d}\lambda}\right) \mathrm{d}\lambda - \psi(1).$$

Thus h_c is likelihood separable.

(iii). Let c be posterior separable, take ϕ^{π^*} as in Definition (8). To shorten notation, we drop the superscript and write ϕ instead of ϕ^{π^*}. As in the proof of Proposition 8, we see that c^* is posterior separable with $\psi = \hat{\phi}_{\pi^*}$. ■

Proof of Proposition 10. "Only if." Define $f : (0, \infty) \to (-\infty, \infty]$ by $f(t) = \psi(t, 1)$. Since ψ finite on $\mathbb{R}^{\{0,1\}}_{++}$, the range of f is included by $\mathbb{R}$. Since ψ is convex, f is convex as well. For every $P \in \mathcal{E}$

$$
\begin{aligned}
h(P) + f(1) &= \int_{\frac{\mathrm{d}P_1}{\mathrm{d}\lambda}>0} \frac{\mathrm{d}P_1}{\mathrm{d}\lambda} \psi\left(\frac{\mathrm{d}P_0/\lambda}{\mathrm{d}P_1/\lambda}, 1\right) \mathrm{d}\lambda + \psi(1,0) \int_{\frac{\mathrm{d}P_0}{\mathrm{d}\lambda} > \frac{\mathrm{d}P_1}{\mathrm{d}\lambda} = 0} \frac{\mathrm{d}P_0}{\mathrm{d}\lambda} \mathrm{d}\lambda \\
&= \int_{\frac{\mathrm{d}P_1}{\mathrm{d}\lambda}>0} \frac{\mathrm{d}P_1}{\mathrm{d}\lambda} f\left(\frac{\mathrm{d}P_0/\lambda}{\mathrm{d}P_1/\lambda}\right) \mathrm{d}\lambda + \left(\lim_{t\to\infty} \frac{f(t)}{t}\right) \int_{\frac{\mathrm{d}P_0}{\mathrm{d}\lambda} > \frac{\mathrm{d}P_1}{\mathrm{d}\lambda} = 0} \frac{\mathrm{d}P_0}{\mathrm{d}\lambda} \mathrm{d}\lambda \\
&= D_f(P_0 \| P_1). \qquad (18)
\end{aligned}
$$

The desired result follows.

"If." Define $\psi : \mathbb{R}^{\{0,1\}}_{+} \to (-\infty, \infty]$ by

$$
\psi(z_0, z_1) = \begin{cases} z_1 f\left(\frac{z_0}{z_1}\right) & \text{if } z_1 > 0 \\ z_1 \lim_{t\to\infty} \frac{f(t)}{t} & \text{if } z_0 > z_1 = 0 \\ 0 & \text{if } z_0 = z_1 = 0. \end{cases}
$$

The function ψ is called the *perspective function* of f (Hiriart-Urruty and Lemaréchal, 2012, Section 2.2). It is a routine exercise to show that ψ is sublinear and lower semicontinuous. Notice also that ψ is finite on $\mathbb{R}^{\{0,1\}}_{++}$. Reasoning as in (18), we see that

$$
h(P) = \int_X \psi\left(\frac{\mathrm{d}P_0}{\mathrm{d}\lambda}, \frac{\mathrm{d}P_1}{\mathrm{d}\lambda}\right) \mathrm{d}\lambda - \psi(1),
$$

which means that h is likelihood separable. ■

Proof of Proposition 11. To prove the result, it is enough to show that $B(\pi, P^\mu) = \mu$ implies $V(P^\mu) = W(\mu)$. Let $\phi \in Cv(\Delta)$ be given by

$$
\phi(p) = \max_{a\in A} \sum_\theta v(a, \theta, c(\mu)) p(\theta).
$$

Because h and c are dual to each other, $h(P^\mu) = c(\mu)$. Thus,

$$
\begin{aligned}
V(P^\mu) &= \int_X \left(\max_{a\in A} \sum_\theta v(a, \theta, c(\mu)) p_x(\theta)\right) \mathrm{d}P_\pi(x) \\
&= \int_X \left(\max_{a\in A} \sum_\theta v(a, \theta, c(\mu)) \pi(\theta) \frac{\mathrm{d}P_\theta(x)}{\mathrm{d}\lambda(x)}\right) \mathrm{d}\lambda(x) = \int_X \hat{\phi}_\pi\left(\frac{\mathrm{d}P(x)}{\mathrm{d}\lambda(x)}\right) \mathrm{d}\lambda(x).
\end{aligned}
$$

It follows from Lemma 9 that

$$\int_X \hat{\phi}_\pi \left(\frac{\mathrm{d}P(x)}{\mathrm{d}\lambda(x)} \right) \mathrm{d}\lambda(x) = \int_\Delta \phi(p) \, \mathrm{d}\mu(p) = W(\mu).$$

We conclude that $V(P^\mu) = W(\mu)$ as desired. ■

Proof of Proposition 12. It is obvious that (ii) implies (i), so we focus on the other implication. Given that the state is binary, we identify Δ and the unit interval $[0,1]$ under the convention that $\pi \in [0,1]$ is the probability that $\theta = 1$.

By (i), we can assume without loss of generality that

$$\lim_{\pi \to 1} c\left((1-\pi)\delta_0 + \pi\delta_1\right) = 0. \tag{19}$$

The random posterior $(1-\pi)\delta_0 + \pi\delta_1$ corresponds to the case in which the agent learns the state perfectly: with probability $1-\pi$, the posterior belief is $p = 0$; with probability π, the posterior belief is $p = 1$.

Let $\mathcal{E}^\infty$ the set of sequential experiments of any length: $\mathcal{E}^\infty = \bigcup_{n=1}^\infty \mathcal{E}^n$. We denote by P^∞ a generic element of $\mathcal{E}^\infty$. For every $n = 1, 2, \ldots, \infty$, let $\hat{\mathcal{E}}^n \subseteq \mathcal{E}^n$ be the set of sequential experiments that perfectly reveals the state:

$$\hat{\mathcal{E}}^n = \{P^n \in \mathcal{E}^n : B(\pi^*, P^n) = (1-\pi^*)\delta_0 + \pi^*\delta_1\}.$$

By hypothesis, c is experimental of order $n = 1, 2, \ldots, \infty$. Thus, for every $\pi \in (0,1)$,

$$c\left((1-\pi)\delta_0 + \pi\delta_1\right) = \inf_{P^n \in \hat{\mathcal{E}}^n} h(\pi, P^n) = \inf_{P^n \in \hat{\mathcal{E}}^n} (1-\pi)h(0, P^n) + \pi h(1, P^n).$$

Claim 1. For every $\epsilon > 0$, there exists $P^n \in \hat{\mathcal{E}}^n$ such that $h(1, P^n) \leq \epsilon$.

Proof of the claim. By (1), for every $\pi \in (0,1)$ sufficiently close 1 there exists $P^{n,\pi} \in \hat{\mathcal{E}}^n$ such that

$$h(\pi, P^{n,\pi}) \leq \frac{\epsilon}{2}.$$

Since $h(\pi, P^{n,\pi}) = (1-\pi)h(0, P^{n,\pi}) + \pi h(1, P^{n,\pi})$ and $h(0, P^{n,\pi}) \geq 0$, we have

$$\pi h(1, P^{n,\pi}) \leq h(\pi, P^{n,\pi}).$$

Thus, we can choose π sufficiently close to one such that

$$h(1, P^{n,\pi}) \leq \frac{h(\pi, P^{n,\pi})}{\pi} \leq \frac{\epsilon}{2\pi} \leq \epsilon.$$

The desired result follows. □

Let $P \in \mathcal{E}$ be an uninformative experiment such that $h(P) = 0$. For $P^n \in \hat{\mathcal{E}}^n$, $\pi \in (0, 1)$, and $\lambda \in (0, \pi)$, define $Q^n \in \mathcal{E}^n$ inductively as follows:

- $Q_{x^0} = P_{x^0}$
- if $\frac{\mathrm{d}Q_1^{i-1}(x^{i-1})}{\mathrm{d}Q_\pi^{i-1}(x^{i-1})} \geq \frac{\lambda}{\pi}$, then $Q_{x^{i-1}} = P_{x^{i-1}}$
- if $\frac{\mathrm{d}Q_1^{i-1}(x^{i-1})}{\mathrm{d}Q_\pi^{i-1}(x^{i-1})} < \frac{\lambda}{\pi}$, then $Q_{x^{t-1}} = P$.

In words, Q^n is the same as P^n with the addition of the following stopping rule: if, after any history of signals x^{i-1}, the posterior belief that the state is one is less than λ, then stop experimenting. Define

$$p_{x^{i-1}} = \pi \frac{\mathrm{d}P_1^{i-1}(x^{i-1})}{\mathrm{d}P_\pi^{i-1}(x^{i-1})} \quad \text{and} \quad q_{x^{i-1}} = \pi \frac{\mathrm{d}Q_1^{i-1}(x^{i-1})}{\mathrm{d}Q_\pi^{i-1}(x^{i-1})}.$$

Given sequential P^n and prior π, $p_{x^{i-1}}$ is the posterior belief that the state is one after history of signals x^{i-1}. Given sequential Q^n and prior π, $q_{x^{i-1}}$ is the posterior belief that the state is one after history of signals x^{i-1}. The random posterior $B(\pi, P^n)$ is the pushforward of P_π^n under $x^n \mapsto p_{x^n}$. The random posterior $B(\pi, Q^n)$ is the pushforward of Q_π^n under $x^n \mapsto q_{x^n}$.

Claim 2. For Q_π^n-almost all x^n, $q_{x^n} \notin [\lambda, 1 - \lambda]$.

Proof of the claim. By construction of Q^n,

$$Q_\pi^n(\{x^n : q_{x^n} \in [\lambda, 1 - \lambda]\}) = P_\pi^n(\{x^n : p_{x^n} \in [\lambda, 1 - \lambda]\}).$$

Since P^n reveals the state perfectly,

$$P_\pi^n(\{x^n : p_{x^n} \in [\lambda, 1 - \lambda]\}) = 0.$$

The desired result follows. □

Claim 3. $\lambda h(\pi, Q^n) \leq h(1, P^n)$

Proof of the claim. It is enough to show, for every i,

$$\lambda \int_{X^{i-1}} h\left(Q_{x^{i-1}}\right) \mathrm{d}Q_\pi^{i-1}(x^{i-1}) \leq \int_{X^{i-1}} h\left(P_{x^{i-1}}\right) \mathrm{d}P_1^{i-1}(x^{i-1}).$$

By construction of Q^n,

$$\begin{aligned}\lambda \int_{X^{i-1}} h\left(Q_{x^{i-1}}\right) \mathrm{d}Q_\pi^{i-1}(x^{i-1}) &= \lambda \int_{\{x^{i-1}: q_{x^{i-1}} \geq \lambda\}} h\left(Q_{x^{i-1}}\right) \mathrm{d}Q_\pi^{i-1}(x^{i-1}). \\ &\leq \pi \int_{\{x^{i-1}: q_{x^{i-1}} \geq \lambda\}} h\left(Q_{x^{i-1}}\right) \mathrm{d}Q_1^{i-1}(x^{i-1}) \\ &= \pi \int_{\{x^{i-1}: p_{x^{i-1}} \geq \lambda\}} h\left(P_{x^{i-1}}\right) \mathrm{d}P_1^{i-1}(x^{i-1}) \\ &\leq \int_{X^{i-1}} h\left(P_{x^{i-1}}\right) \mathrm{d}P_1^{i-1}(x^{i-1}).\end{aligned}$$

The desired result follows. □

Claim 4. For every $\pi \in (0,1)$, there exists a net $(\mu^\epsilon)_{\epsilon>0}$ in Δ^2_π such that, as $\epsilon \to 0$, $\mu^\epsilon \to (1-\pi)\delta_0 + \pi\delta_1$ and $c(\mu^\epsilon) \to 0$.

Proof of the claim. Fix $\pi \in (0,1)$. For every $\epsilon > 0$, choose $P^{n,\epsilon} \in \hat{\mathcal{E}}^n$ such that $h(1, P^{n,\epsilon}) \leq \epsilon^2$ (see Claim 1). Setting $\lambda = \epsilon$, we can find $Q^{n,\epsilon}$ that satisfy the following conditions:

- For $Q^{n,\epsilon}_\pi$-almost all x^n, $q^\epsilon_{x^n} \notin [\epsilon, 1-\epsilon]$ (see Claim 2).
- $h(\pi, Q^{n,\epsilon}) \leq \frac{1}{\lambda} h(1, P^{n,\epsilon}) = \epsilon$ (see Claim 3).

Define $\mu^\epsilon = B(\pi, Q^{n,\epsilon})$. Then $c(\mu^\epsilon) \to 0$ as $\epsilon \to 0$. Assume without loss of generality that, for some $\mu \in \Delta^2_+$, $\mu^\epsilon \to \mu$ as $\epsilon \to 0$. For every $\eta \in (0, 1/2)$,

$$\mu((\eta, 1-\eta)) \leq \liminf_{\epsilon \to 0} \mu^\epsilon((\eta, 1-\eta)) \leq \liminf_{\epsilon \to 0} \mu^\epsilon([\eta, 1-\eta]) = 0.$$

By σ-continuity, $\mu((\eta, 1-\eta)) \to \mu((0,1))$ as $\eta \to 0$. Thus $\mu((0,1)) = 0$. We conclude that $\mu = (1-\pi)\delta_0 + \pi\delta_1$. The desired result follows. □

We are ready to conclude the proof. Take any $\mu \in \Delta^2_+$. Choose $(\mu^\epsilon)_{\epsilon>0}$ in $\Delta^2_{\bar{\mu}}$ such that, as $\epsilon \to 0$, $\mu^\epsilon \to (1-\bar{\mu})\delta_0 + \bar{\mu}\delta_1$ and $c(\mu^\epsilon) \to 0$ (see Claim 4). Then

$$c(\mu) \leq c((1-\bar{\mu})\delta_0 + \bar{\mu}\delta_1) \leq \lim_{\epsilon \to 0} c(\mu^\epsilon) = 0$$

where we use that c is monotone in the convex order and lower semicontinuous. We conclude that $c(\mu) = 0$, as desired. ■

Proof of Proposition 13. Define $\pi_\alpha = \alpha\pi + (1-\alpha)\rho$. If $B(\pi_\alpha, P^n) = B(\pi_\alpha, P)$, then $B(\pi, P^n) = B(\pi, P)$ and $B(\rho, P^n) = B(\rho, P)$ by Lemma 8. In addition,

$$h^n(\pi_\alpha, P^n) = \alpha h^n(\pi, P^n) + (1-\alpha) h^n(\rho, P^n).$$

Therefore we obtain

$$h^n(\pi_\alpha, P^n) \geq \alpha c_{h^n}(B(\pi, P)) + (1-\alpha)c_{h^n}(B(\rho, P)).$$

Since the choice of P^n is arbitrary, we conclude that

$$c_{h^n}(B(\pi_\alpha, P)) \geq \alpha c_{h^n}(B(\pi, P)) + (1-\alpha)c_{h^n}(B(\rho, P))$$

as desired. ■

Proof of Proposition 14. Let $\phi \in Cv(\Delta)$ be defined by

$$\phi(p) = \max_{a\in A} \sum_{\theta} v(a, \theta)p(\theta).$$

We can rewrite the information acquisition problem as

$$V(\pi) = \max_{\mu\in\Delta^2_\pi} \int_\Delta \phi(p)\, \mathrm{d}\mu(p) - c(\mu).$$

Being $\mathcal{E}$ rich enough to induce any random posterior,

$$V(\pi) = \max_{P\in\mathcal{E}} \int_\Delta \phi\, \mathrm{d}B(\pi, P) - c(B(\pi, P)).$$

The quantity $\int_\Delta \phi\, \mathrm{d}B(\pi, P)$ is a convex function of $\pi \in \Delta_+$. Indeed, for $\pi, \rho \in \Delta_+$ and $\alpha \in [0, 1]$, we have by Lemma 10 that

$$\alpha B(\pi, P) + (1-\alpha)B(\rho, P) \succeq_{cv} B(\alpha\pi + (1-\alpha)\rho, P).$$

We obtain that

$$\alpha \int_\Delta \phi\, \mathrm{d}B(\pi, P) + (1-\alpha) \int_\Delta \phi\, \mathrm{d}B(\rho, P) \geq \int_\Delta \phi\, \mathrm{d}B(\alpha\pi + (1-\alpha)\rho, P).$$

As a result, because $c(B(\pi, P))$ is a concave function of $\pi \in \Delta_+$, then quantity

$$\int_\Delta \phi\, \mathrm{d}B(\pi, P) - c(B(\pi, P))$$

is a convex function of $\pi \in \Delta_+$. Since the supremum of convex functions is a convex function, we conclude that $V(\pi)$ is a convex function of $\pi \in \Delta_+$. ■

C Priors with partial support

In the main text we focus on full-support priors. The extension to arbitrary priors presents no difficulties, as we sketch in this section. Recall that Δ^2 is the set of all random posteriors μ, with arbitrary barycenters $\bar{\mu} \in \Delta$. As a consequence of Lemma 7, we have

$$\Delta^2 = \{B(\pi, P) : \pi \in \Delta, P \in \mathcal{E}\}.$$

A cost function $c : \Delta^2 \to [0, \infty]$ is *experimental* if there is $h : \mathcal{E} \to [0, \infty]$ such that

$$c(\mu) = \inf \{h(P) : B(\bar{\mu}, P) = \mu\} \qquad \forall \mu \in \Delta^2. \tag{20}$$

We write $c = c_h$ and say that c is induced by h.

To characterize experimental cost functions on the full domain Δ^2, we introduce a sub-order of $\succeq_{ex}$.

Definition 13. The *sub-experimental order* $\succeq_{sx}$ is a binary relation on Δ^2 defined by $\mu \succeq_{sx} \nu$ if, for every $P \in \mathcal{E}$ such that $B(\bar{\mu}, P) = \mu$, there is $Q \in \mathcal{E}$ such that $P \sim_b Q$ and $B(\bar{\nu}, Q) = \nu$.

Thus, $\mu \succeq_{sx} \nu$ if and only if

$$\{P : B(\bar{\mu}, P) = \mu\} \subseteq \{P : B(\bar{\nu}, P) = \nu\}. \tag{21}$$

Moreover, $\mu \succeq_{sx} \nu$ implies $\mu \succeq_{ex} \nu$, but the converse may not hold. In view of Lemma 1-(ii), $\mu \sim_{sx} \nu$ if and only if $\mu \sim_{ex} \nu$. With a slight about of notation, denote by $\mathcal{C}$ the class of cost function $c : \Delta^2 \to [0, \infty]$ such that $c(\delta_\pi)$ for all $\pi \in \Delta$.

Proposition 15. *(i). A cost function $c \in \mathcal{C}$ is experimental if and only if is monotone in the sub-experimental order and, for every $\mu \in \Delta^2$, there is a sequence (ν_n) in Δ^2_+ such that $c(\nu_n) \to c(\mu)$ and, for every n, $\nu_n \succeq_{sx} \mu$. Moreover, c is induced by a unique $h_c \in \mathcal{H}$ that is invariant under $\sim_b$, given by*

$$h_c(P) = c(B(\pi^*, P)). \tag{22}$$

(ii). If $h \in \mathcal{H}$ is monotone in the Blackwell order, then c_h is monotone in the convex order and

$$c_h(\mu) = \inf\{h(P) : B(\bar{\mu}, P) \succeq_{cv} \mu\} \qquad \forall \mu \in \Delta^2.$$

(iii). A cost function $c \in \mathcal{C}$ is monotone in the convex order and experimental if and only if is monotone in the experimental order and, for every $\mu \in \Delta^2$, there is a sequence

(ν_n) in Δ^2_+ such that $\nu_n \succeq_{ex} \mu$ for every n, and $c(\nu_n) \to c(\mu)$. Moreover, h_c is Blackwell monotone.

A function $c : \Delta^2_+ \to [0, \infty]$ is monotone in $\succeq_{sx}$ if and only if is invariant under $\sim_{ex}$. Indeed, if $\mu \succeq_{sx} \nu$ and $\mu, \nu \in \Delta^2_+$, then $\mu \sim_{ex} \nu$; conversely, if $\mu \sim_{ex} \nu$, then $\mu \sim_{sx} \nu$. Proposition 15 therefore generalizes Theorem 1.

Proof. (i). "Only if." Assume that c is experimental. It follows from (21) that c is monotone in the sub-experimental order. To see that also (ii) is satisfied, take $h : \mathcal{E} \to [0, \infty]$ such that $c = c_h$. For every $\mu \in \Delta^2$, there is a sequence (P_n) in $\mathcal{E}$ such that $B(\bar{\mu}, P_n) = \mu$ for every n, and $h(P_n) \to c(\mu)$. Define $\nu_n = B(\pi^*, P_n)$.

We claim that $\nu_n \succeq_{sx} \mu$. Indeed, take Q such that $B(\pi^*, Q) = \nu_n$. By Lemma 8 we have $Q \sim_b P_n$, thus $B(\bar{\mu}, Q) = B(\bar{\mu}, P_n) = \mu$. By (21) we obtain that $\nu_n \succeq_{sx} \mu$ as desired. Now, being $c = c_h$, $h(P_n) \geq c(\nu_n)$. Moreover, being c monotone in the sub-experimental order (as shown above), $c(\nu_n) \geq c(\mu_n)$. Thus $h(P_n) \to c(\mu)$ implies $c(\nu_n) \to c(\nu)$.

"If." Define $h : \mathcal{E} \to [0, \infty]$ by

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

By hypothesis, c is monotone in $\succeq_{sx}$. Thus, $\mu \sim_{ex} \nu$ (which is equivalent to $\mu \sim_{sx} \nu$) implies $c(\mu) = c(\nu)$. By Theorem 1, $c(\mu) = c_h(\mu)$ for all $\mu \in \Delta^2_+$.

We claim that, for all $\mu \in \Delta^2$, $c_h(\mu) \geq c(\mu)$. To see this, take an experiment P such that $B(\bar{\mu}, P) = \mu$. By construction, $h(P) = c(B(\pi^*, P))$. By Lemma 8 and (21), $B(\pi^*, P) \succeq_{sx} \mu$. Being c monotone in the sub-experimental order (by hypothesis), $h(P) = c(B(\pi^*, P)) \geq c(\mu)$. Since the choice of P was arbitrary, $c_h(\mu) \geq c(\mu)$.

We now argue that $c_h(\mu) = c(\mu)$ and, therefore, that c is experimental. By contradiction, suppose that $c(\mu) \neq c_h(\mu)$. As shown above $c_h(\mu) \geq c(\mu)$, thus it must be that $c_h(\mu) > c(\mu)$. By hypothesis, there is a sequence (ν_n) in Δ^2_+ such that $c(\nu_n) \to c(\mu)$ and, for every n, $\nu_n \succeq_{sx} \mu$. As shown above, $c(\nu_n) = c_h(\nu_n)$ for every n, thus $c_h(\nu_n) \to c(\mu)$. Eventually $c_h(\mu) > c_h(\nu_n)$, being that $c_h(\mu) > c(\mu)$. But c_h is monotone in the sub-experimental order, which implies $c_h(\nu_n) \geq c_h(\mu)$ for every n: contradiction.

(ii). Let $h : \mathcal{E} \to [0, \infty]$ be monotone in the Blackwell order. By definition,

$$c_h(\mu) \geq \inf\{h(P) : B(\bar{\mu}, P) \succeq_{cv} \mu\}.$$

To prove the opposite inequality, take $P, P^\mu \in \mathcal{E}$ such that $B(\bar{\mu}, P) \succeq_{cv} \mu = B(\bar{\mu}, P^\mu)$. By Lemma 8, there is a stochastic kernel K such that, for all $\theta \in \operatorname{supp} \bar{\mu}$, $P^\mu_\theta = K P_\theta$. Define $Q \in \mathcal{E}$ by, for all θ, $Q_\theta = K P_\theta$. Then $P \succeq_b Q$ and $\mu = B(\bar{\mu}, Q)$. Since h is Blackwell monotone, $h(P) \geq h(Q)$. Since the choice of P was arbitrary, we conclude that

$$c_h(\mu) \leq \inf\{h(P) : B(\bar{\mu}, P) \succeq_{cv} \mu\}.$$

Overall, we obtain

$$c_h(\mu) = \inf\{h(P) : B(\bar{\mu}, P) \succeq_{cv} \mu\}.$$

By Lemma 1-(i), c_h is monotone in the experimental order. By 1-(iii), c_h is monotone in the convex order.

(iii). "Only if." Assume that c is monotone in the convex order and experimental. By (i), $c = c_{h_c}$. Since c is monotone in the convex order, h_c is monotone in the Blackwell order (Lemma 8). Thus, by (ii), c is monotone in the experimental order. In addition, again by (i), for every $\mu \in \Delta^2$ there is a sequence (ν_n) in Δ^2_+ such that $c(\nu_n) \to c(\mu)$ and, for every n, $\nu_n \succeq_{sx} \mu$, which implies $\nu_n \succeq_{ex} \mu$.

"If." Define $h : \mathcal{E} \to [0, \infty]$ by

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

By hypothesis, c is monotone in $\succeq_{ex}$. Thus, by Theorem 1, $c(\mu) = c_h(\mu)$ for all $\mu \in \Delta^2_+$. Moreover, h is Blackwell monotone.

We claim that, for all $\mu \in \Delta^2$, $c_h(\mu) \geq c(\mu)$. To see this, take an experiment P such that $B(\bar{\mu}, P) = \mu$. By construction, $h(P) = c(B(\pi^*, P))$. By Lemmas 8 and 1, $B(\pi^*, P) \succeq_{ex} \mu$. Being c monotone in the experimental order (by hypothesis), $h(P) = c(B(\pi^*, P)) \geq c(\mu)$. Since the choice of P was arbitrary, $c_h(\mu) \geq c(\mu)$.

We now argue that $c_h(\mu) = c(\mu)$ and, therefore, that c is experimental. By contradiction, suppose that $c(\mu) \neq c_h(\mu)$. As shown above $c_h(\mu) \geq c(\mu)$, thus it must be that $c_h(\mu) > c(\mu)$. By hypothesis, there is a sequence (ν_n) in Δ^2_+ such that $c(\nu_n) \to c(\mu)$ and, for every n, $\nu_n \succeq_{ex} \mu$. As shown above, $c(\nu_n) = c_h(\nu_n)$ for every n, thus $c_h(\nu_n) \to c(\mu)$. Eventually $c_h(\mu) > c_h(\nu_n)$, being that $c_h(\mu) > c(\mu)$. But c_h is monotone in the experimental order, which implies $c_h(\nu_n) \geq c_h(\mu)$ for every n: contradiction. □

The duality map and the regularization map extend to the case of priors with partial support. Every function $h \in \mathcal{H}$ that is invariant under $\sim_b$ induces a unique experimental cost function $c_h : \Delta^2 \to [0, \infty]$ defined by (20). Conversely, every experimental cost function $c : \Delta^2 \to [0, \infty]$ is induced by a unique function $h_c : \mathcal{E} \to [0, \infty]$ that is invariant under $\sim_b$ defined by (22). If $c \in \mathcal{C}$ is any cost function, then the function $c^* : \Delta^2 \to [0, \infty]$ defined by

$$c^*(\mu) = \inf\{c(\nu) : \bar{\nu} = \pi^* \text{ and } \nu \succeq_{sx} \mu\}$$

is experimental. In particular, $c = c^*$ if and only if c is experimental.

Next we illustrate the regularization of rational inattention for general priors.

Example 13. Fix $\mu \in \Delta^2$ and denote by $S \subseteq \Theta$ the support of $\bar{\mu}$. We claim that

$$c_R^*(\mu) = \sum_{\theta \in S} \pi^*(\theta) D_{KL}\left(P_\theta^\mu \| P_{\pi_S^*}^\mu\right) \tag{23}$$

where P^μ satisfies $B(\bar{\mu}, P^\mu) = \mu$, and $\pi_S^* \in \Delta$ is the uniform distribution over S.

To verify (23), take $P \in \mathcal{E}$ such that $B(\bar{\mu}, P) = \mu$. Let $Q, Q^c \in \mathcal{E}$ be given by

$$Q_\theta = \begin{cases} P_\theta & \text{if } \theta \in S \\ P_{\pi_S^*} & \text{otherwise} \end{cases} \qquad \text{and} \qquad Q_\theta^c = \begin{cases} P_\theta & \text{if } \theta \notin S \\ P_{\pi_{S^c}^*} & \text{otherwise.} \end{cases}$$

Since $B(\bar{\mu}, P) = \mu$ and $P_\theta = Q_\theta$ for all $\theta \in S$, we have $B(\bar{\mu}, Q) = \mu$. Simple algebra shows that

$$h_R(P) \geq h_R(Q) + h_R(Q^c) \geq h_R(Q) = \sum_{\theta \in S} \pi^*(\theta) D_{KL}\left(P_\theta \| P_{\pi_S^*}\right).$$

We deduce that

$$c_R^*(\mu) = \inf\{h_R(P) : B(P, \bar{\mu}) = \mu\} = \inf\left\{\sum_{\theta \in S} \pi^*(\theta) D_{KL}\left(P_\theta \| P_{\pi_S^*}\right) : B(P, \bar{\mu}) = \mu\right\}.$$

The condition $B(\bar{\mu}, P) = \mu = B(\bar{\mu}, P^\mu)$ implies that

$$\sum_{\theta \in S} \pi^*(\theta) D_{KL}\left(P_\theta \| P_{\pi_S^*}\right) = \sum_{\theta \in S} \pi^*(\theta) D_{KL}\left(P_\theta^\mu \| P_{\pi_S^*}^\mu\right).$$

It follows that (23) holds. ▲

References

S. M. Ali and S. D. Silvey. A general class of coefficients of divergence of one distribution from another. *Journal of the Royal Statistical Society, Series B*, 28:131–142, 1966.

C. D. Aliprantis and K. C. Border. *Infinite Dimensional Analysis.* Springer, 3rd edition, 2006.

R. Alonso and O. Camara. Bayesian persuasion with heterogeneous priors. *Journal of Economic Theory*, 165:672–706, 2016.

K. J. Arrow, D. Blackwell, and M. A. Girshick. Bayes and minimax solutions of sequential decision problems. *Econometrica*, 17:213–244, 1949.

D. Bergemann and J. Välimäki. Information acquisition and efficient mechanism design. *Econometrica*, 70:1007–1033, 2002.

D. Blackwell. Comparison of experiments. In *Proc. 2nd Berkeley Symp. on Math. Stat. and Probab.*, pages 93–102. University of California Press, 1951.

D. Blackwell and M. A. Girshick. *Theory of Games and Statistical Decisions*. John Wiley & Sons, 1954.

A. W. Bloedel and W. Zhong. The cost of optimally acquired information. Technical report, 2020.

S. Board and J. Lu. Competitive information disclosure in search markets. *Journal of Political Economy*, 126:1965–2010, 2018.

H. F. Bohnenblust, L. S. Shapley, and S. Sherman. Reconnaissance in game theory. 1949.

A. Cabrales, O. Gossner, and R. Serrano. Entropy and the value of information for investors. *American Economic Review*, 103:360–377, 2013.

A. Caplin and M. Dean. Behavioral implications of rational inattention with Shannon entropy. 2013.

A. Caplin and M. Dean. Revealed preference, rational inattention, and costly information acquisition. *American Economic Review*, 105:2183–2203, 2015.

A. Caplin, M. Dean, and J. Leahy. Rationally inattentive behavior: Characterizing and generalizing Shannon entropy. 2021.

T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley & Sons, 2nd edition, 2012.

I. Csiszár. Eine informationstheoretische ungleichung und ihre anwendung auf beweis der ergodizitaet von markoffschen ketten. *Magyer Tud. Akad. Mat. Kutato Int. Koezl.*, 8: 85–108, 1963.

H. De Oliveira, T. Denti, M. Mihm, and K. Ozbek. Rationally inattentive preferences and hidden information costs. *Theoretical Economics*, 12:621–654, 2017.

J. Geanakoplos, D. Pearce, and E. Stacchetti. Psychological games and sequential rationality. *Games and Economic Behavior*, 1:60–79, 1989.

M. Gentzkow and E. Kamenica. Costly persuasion. *American Economic Review P&P*, 104: 457–462, 2014.

B. Hébert and M. Woodford. Rational inattention when decisions take time. 2019.

J.-B. Hiriart-Urruty and C. Lemaréchal. *Fundamentals of Convex Analysis*. Springer, 2012.

L. Le Cam. Comparison of experiments: A short review. *Lecture Notes–Monograph Series*, 30:127–138, 1996.

F. Liese and I. Vajda. On divergences and informations in statistics and information theory. *IEEE Transactions on Information Theory*, 52:4394–4412, 2006.

B. Maćkowiak, F. Matějka, and M. Wiederholt. Rational inattention: A review. *Journal of Economic Literature (forthcoming)*, 2021.

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

J. Mensch. Cardinal representations of information. 2018.

J. Miao and H. Xing. Dynamic discrete choice under rational inattention. 2020.

S. Morris and P. Strack. The Wald problem and the equivalence of sequential sampling and static information costs. 2019.

K. Nimark and S. Sundaresan. Inattention and belief polarization. *Journal of Economic Theory*, 180:203–228, 2019.

N. Persico. Information acquisition in auctions. *Econometrica*, 68:135–148, 2000.

L. Pomatto, P. Strack, and O. Tamuz. The cost of information. 2020.

D. Ravid. Ultimatum bargaining with rational inattention. *American Economic Review*, 110:2948–2963, 2020.

Doron Ravid, Anne-Katrin Roesler, and Balazs Szentes. Learning before trading: on the inefficiency of ignoring free information. 2021.

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

Christopher A Sims. Rational inattention and monetary economics. In *Handbook of monetary economics*, volume 3, pages 155–181. Elsevier, 2010.

E. Torgersen. *Comparison of Statistical Experiments*. Cambridge University Press, 1991.

A. Wald. Foundations of a general theory of sequential decision functions. *Econometrica*, 15:279–313, 1947.

A. Wald. *Statistical Decision Functions.* John Wiley & Sons, 1950.