Posterior Separable Cost of Information: Online Appendix

Tommaso Denti

B An equivalent convex program

In many applications of rational inattention, a choice from among distributions of posteriors $\mu \in \mathcal{M}_\pi$ represents the acquisition of information. Studying such choice may not be straightforward since distributions of posteriors are infinite dimensional objects. A common approach is to re-formulate the information acquisition problem as a choice from among stochastic kernels $\sigma_F \in \mathcal{S}_F$, which is a more tractable program. For example, in the leading case of the entropy cost, Matejka and McKay (2015) study the program

$$\max_{\sigma_F \in \mathcal{S}_F} \sum_{f,\theta} f(\theta)\sigma_F(f|\theta)\pi(\theta) - \sum_{f,\theta} \sigma_F(f|\theta)\pi(\theta) \log \frac{\sigma_F(f|\theta)}{\sigma_F(f)}. \tag{34}$$

The program has two essential features: it is finite dimensional and it is convex. This allows Matejka and McKay (2015) to use standard solution methods based on Lagrange multipliers.

In this section we describe a generalization of (34) to all posterior separable costs. The generalization we propose maintain the two essential features of (34): finite dimensionality and convexity.

Let $H : \Delta(\Theta) \to \mathbb{R}$ be a concave measure of uncertainty. We generalize (34) by building on the superlinear extension $\hat{H} : \mathbb{R}^\Theta_+ \to \mathbb{R}$ of the function H. Letting $\Theta = \{1, \dots n\}$ and $\zeta = (\zeta_1, \dots, \zeta_n) \in \mathbb{R}^\Theta_+$, the function $\hat{H}$ is given by

$$\hat{H}(\zeta_1, \dots, \zeta_n) = \begin{cases} \left(\sum_{i=1}^n \zeta_i\right) H\left(\frac{\zeta_1}{\sum_{i=1}^n \zeta_i}, \dots, \frac{\zeta_n}{\sum_{i=1}^n \zeta_i}\right) & \text{if } \sum_{i=1}^n \zeta_i > 0, \\ 0 & \text{otherwise.} \end{cases}$$

The function $\hat{H}$ extends H: for all $p \in \Delta(\Theta)$, $H(p) = \hat{H}(p)$. Since H is concave and upper semicontinuous, the function $\hat{H}$ is superlinear and upper semicontinuous. "Superlinear" means that for all $\alpha, \beta \in \mathbb{R}_+$ and $\zeta, \eta \in \mathbb{R}^\Theta_+$,

$$\hat{H}(\alpha\zeta + \beta\eta) \geq \alpha\hat{H}(\zeta) + \beta\hat{H}(\eta).$$

It is easy to see that $\hat{H}$ is the only superlinear extension of H from $\Delta(\Theta)$ to $\mathbb{R}^\Theta_+$.

The next lemma generalizes (34). To state the result, given $\sigma_F \in \mathcal{S}_F$ and $f \in F$, we denote by $\sigma_{(F,f)} \in \mathbb{R}_+^\Theta$ the vector $(\sigma_F(f|\theta))_{\theta\in\Theta}$. We also denote by $\sigma_{(F,f)}\pi$ the pointwise product of $\sigma_{(F,f)}$ and π:

$$\sigma_{(F,f)}\pi = (\sigma_F(f|\theta)\pi(\theta))_{\theta\in\Theta}.$$

Lemma 9. *Suppose the cost of information is posterior separable with concave uncertainty measure H. If $(\mu, d_F) \in \mathcal{M}_\pi \times \mathcal{D}_F$ is an optimal solution of (1), then $\sigma_{(\mu,d_F)}$ is an optimal solution of*

$$\max_{\sigma_F \in \mathcal{S}_F} \sum_{f,\theta} f(\theta)\sigma_F(f|\theta)\pi(\theta) - \left(\hat{H}(\pi) - \sum_f \hat{H}\left(\sigma_{(F,f)}\pi\right)\right). \tag{35}$$

Conversely, if $\sigma_F \in \mathcal{S}_F$ is an optimal solution of (35), then $(\mu_{\sigma_F}, d_{\sigma_F})$ is an optimal solution of (1).

Proof. Let $(\mu, d_F) \in \mathcal{M}_\pi \times \mathcal{D}_F$ and $\sigma_F \in \mathcal{S}_F$. Simple algebra shows that

$$\sum_{f,\theta} f(\theta)\sigma_F(f|\theta)\pi(\theta) = \int_{\Delta(\Theta)} \left(\sum_f (f\cdot p) d_{\sigma_F}(f|p)\right) \mathrm{d}\mu_{\sigma_F}(p), \tag{36}$$

$$\sum_f \hat{H}\left(\sigma_{(F,f)}\pi\right) = \int_{\Delta(\Theta)} H(p)\,\mathrm{d}\mu_{\sigma_F}(p). \tag{37}$$

In addition, by Lemma 1, if $\sigma_F = \sigma_{(\mu,d_F)}$ then

$$\int_{\Delta(\Theta)} \left(\sum_f (f\cdot p)\, d_{\sigma_F}(f|p)\right) \mathrm{d}\mu_{\sigma_F}(p) = \int_{\Delta(\Theta)} \left(\sum_f (f\cdot p) d_F(f|p)\right) \mathrm{d}\mu(p), \tag{38}$$

$$\int_{\Delta(\Theta)} H(p)\,\mathrm{d}\mu_{\sigma_F}(p) \geq \int_{\Delta(\Theta)} H(p)\,\mathrm{d}\mu(p). \tag{39}$$

Now, suppose that (μ, d_F) is an optimal solution of (1). Set $\sigma_F = \sigma_{(\mu,d_F)}$. By (38) and (39), also $(\mu_{\sigma_F}, d_{\sigma_F})$ is an optimal solution of (1). Therefore, by (36) and (37), σ_F is an optimal solution of (35).

Conversely, suppose that σ_F is an optimal solution of (35). By (36) and (37), the pair $(\mu_{\sigma_F}, d_{\sigma_F})$ maximizes

$$\int_{\Delta(\Theta)} \left(\sum_f (f\cdot p) d_F(f|p)\right) \mathrm{d}\mu(p) - \left(H(\pi) - \int_{\Delta(\Theta)} H(p)\,\mathrm{d}\mu(p)\right)$$

overall all $(\mu, d_F) \in \mathcal{M}_\pi \times \mathcal{D}_F$ such that $(\mu, d_F) = (\mu_{\tau_F}, d_{\tau_F})$ for some $\tau_F \in \mathcal{S}_F$. Therefore, by (38) and (39), the pair $(\mu_{\sigma_F}, d_{\sigma_F})$ is an optimal solution of (1). $\square$

The program (35) generalizes (34); importantly, it preserves finite dimensionality and convexity. If H is entropy, then

$$\sum_{f,\theta} \sigma_F(f|\theta)\pi(\theta) \log \frac{\sigma_F(f|\theta)}{\sigma_F(f)} = \hat{H}(\pi) - \sum_f \hat{H}\left(\sigma_{(F,f)}\pi\right).$$

As in (34), the choice variable is σ_F, a finite dimensional object. Since the function $\hat{H}$ is superlinear—hence, in particular, concave—the program (35) is convex.

Next we use Lagrange multipliers to derive (necessary and sufficient) first order conditions. To state the result, let $\partial\hat{H}(\zeta)$ be the *superdifferential* of $\hat{H}$ at $\zeta \in \mathbb{R}^\Theta_+$. Recall that $\partial\hat{H}(\zeta)$ is the set of all $\zeta^* \in \mathbb{R}^\Theta$ such that for all $\eta \in \mathbb{R}^\Theta_+$,

$$\hat{H}(\zeta) - \zeta \cdot \zeta^* \geq \hat{H}(\eta) - \eta \cdot \zeta^*.$$

Lemma 10. *A stochastic kernel $\sigma_F \in \mathcal{S}_F$ is an optimal solution of (35) if and only if there exists a Lagrange multiplier $\lambda_F \in \mathbb{R}^\Theta$ such that for all $f \in F$,*

$$-f - \lambda_F \in \partial\hat{H}\left(\sigma_{(F,f)}\pi\right). \tag{40}$$

Proof. By a change of variables, σ_F is an optimal solution of (35) if and only if $\left(\sigma_{(F,f)}\pi\right)_{f\in F}$ is an optimal solution of

$$\max_{(\zeta_f)_{f\in F}} \sum_f f \cdot \zeta_f + \sum_f \hat{H}(\zeta_f) \qquad \text{subject to} \quad \sum_f \zeta_f = \pi. \tag{41}$$

Since $\hat{H}$ is concave and upper semicontinuous, it follows from standard convex programming (e.g., Rockafellar 1970, Theorem 28.2) that $\left(\sigma_{(F,f)}\pi\right)_{f\in F}$ is an optimal solution of (41) if and only there is a Lagrange multiplier $\lambda_F \in \mathbb{R}^\Theta$ such that $\left(\sigma_{(F,f)}\pi\right)_{f\in F}$ is an optimal solution of

$$\max_{(\zeta_f)_{f\in F}} \sum_f (f + \lambda_F) \cdot \zeta_f + \sum_f \hat{H}(\zeta_f). \tag{42}$$

By separability of the objective function, $\left(\sigma_{(F,f)}\pi\right)_{f\in F}$ is an optimal solution of (42) if and only if for every $f \in F$,

$$\sigma_{(F,f)}\pi \in \arg\max_{\zeta\in\mathbb{R}^\Theta_+} (f + \lambda_F) \cdot \zeta + \hat{H}(\zeta),$$

which is the same as saying that $-f - \lambda_F$ is an element of $\partial\hat{H}\left(\sigma_{(F,f)}\pi\right)$. □

Lemmas 9 and 10 can be of interest for applications beyond the scope of this paper. The solution method based on Lagrange multipliers complements the concavification technique originally proposed by Caplin and Dean (2013). The concavification technique is based on the observation that when cost of information is posterior separable, the information acquisition problem can be formulated as a concavification problem:

$$\max_{\mu \in \mathcal{M}_\pi} \int_{\Delta(\Theta)} (\phi_F(p) + (H(\pi) - H(p)))\, \mathrm{d}\mu(p).$$

The function $p \mapsto \phi_F(p) + (H(\pi) - H(p))$ is the target of the concavification.

Caplin, Dean and Leahy (2017) also develops a Lagrangian approach to optimal information acquisition when costs are posterior separable; the techniques they develop can be seen as complementary to ours. The use of the superlinear extension of H is peculiar to this paper.

The program (35) reveals a connection between posterior separable costs and *perturbed utility functions* (Fudenberg, Iijima and Strzalecki, 2015). Perturbed utility functions are a model of stochastic choice where the agent's objective function is the sum of expected utility and a non-linear perturbation. To illustrate, suppose that Θ is a singleton, so that an act f is simply a real number. Let $c : [0,1] \to \mathbb{R}$ be a strictly concave function. In Fudenberg, Iijima and Strzalecki (2015), the agent's behavior is the outcome of the optimization

$$\max_{\sigma_F \in \mathcal{S}_F} \sum_f f \cdot \sigma_F(f) + \sum_f c\left(\sigma_F(f)\right). \tag{43}$$

The function c is a perturbation of the agent's utility.

The main similarity between (35) and (43) is that both $\hat{H}$ and c are concave: the agent is rewarded for randomizing. The main difference between (35) and (43) is that $\hat{H}$ is *positively homogenous* (of degree one), while c is *strictly concave*. Since $\hat{H}$ is positively homogenous, (35) does not reward the agent for randomizing independently of the state. Since c is strictly concave, (43) rewards the agent also for noisy behavior. This distinguishes a model of stochastic choice based on incomplete information (e.g., this paper) from a model of stochastic choice based on trembles (e.g., Fudenberg, Iijima and Strzalecki, 2015).

C Computational complexity of Axiom 3

Despite its rich structure, Axiom 3 has reasonably low computational complexity. In practice, analysts can verify whether or not a dataset satisfies Axiom 3 by solving an equivalent linear program, which we formalize next.

Proposition 5. *Let $F_1, \ldots, F_n$ be a finite sequence in $\mathcal{F}$. Axiom 3 is satisfied if and only* *if the following linear program has non-negative value:*

$$\begin{aligned}
\textit{minimize} \quad & \sum_{i=1}^{n} \sum_{f_i \in \operatorname{supp}(\sigma_{F_i})} x_{(F_i,f_i)} \phi_{F_i}\left(p_{(F_i,f_i)}\right) - \sum_{i,j=1}^{n} \sum_{f_j \in \operatorname{supp}(\sigma_{F_j})} y_{(F_i,F_j,f_j)} \phi_{F_i}\left(p_{(F_j,f_j)}\right) \\
\textit{subject to} \quad & x_{(F_i,f_i)} \geq 0 \textit{ and } y_{(F_i,F_j,f_j)} \geq 0, \qquad i,j = 1, \ldots n, f_i \in F_i, f_j \in F_j, \\
& \sum_{i=1}^{n} \sum_{f_i \in \operatorname{supp}(\sigma_{F_i})} x_{(F_i,f_i)} = 1, \qquad (44) \\
& \sum_{f_i \in \operatorname{supp}(\sigma_{F_i})} x_{(F_i,f_i)} p_{(F_i,f_i)} = \sum_{j=1}^{n} \sum_{f_j \in \operatorname{supp}(\sigma_{F_j})} y_{(F_i,F_j,f_j)} p_{(F_j,f_j)}, \qquad i = 1, \ldots, n, \\
& \qquad (45) \\
& x_{(F_j,f_j)} = \sum_{i=1}^{n} y_{(F_i,F_j,f_j)}, \qquad j = 1, \ldots, n, f_j \in \operatorname{supp}(\sigma_{F_j}). \qquad (46)
\end{aligned}$$

To see the relation between the linear program and Axiom 3, let $\mu_i, \nu_i \in \mathcal{M}$ and $\alpha_i \geq 0$ as in Axiom 3. For simplicity, assume that the μ_{F_i} have disjoint supports. Define

$$x_{(F_i,f_i)} = \alpha_i \mu_i \left(p_{(F_i,f_i)}\right) \qquad \text{and} \qquad y_{(F_i,F_j,f_j)} = \alpha_i \nu_i \left(p_{(F_j,f_j)}\right).$$

It is easy to check that the constraints of the linear program are satisfied. In particular, $\bar{\mu}_i = \bar{\nu}_i$ implies (45) and $\sum_i \alpha_i \mu_i = \sum_i \alpha_i \nu_i$ implies (46). The objective function is equal to

$$\sum_{i=1}^{n} \alpha_i \int \phi_{F_i} \, \mathrm{d}\mu_i - \sum_{i=1}^{n} \alpha_i \int \phi_{F_i} \, \mathrm{d}\nu_i.$$

With finite datasets, one can simply take $\{F_1, \ldots, F_n\} = \mathcal{F}$. Thus, for finite datasets, verifying Axiom 3 is equivalent to solving *one* linear program. The result provides a recipe to test Axiom 3 in practice.

Proof of Proposition 5. To simplify notation, assume that $\operatorname{supp}(\sigma_{F_i}) = F_i$ for all $i = 1, \ldots, n$. The more general case easily follows.

"If." Assume the linear program has non-negative value. For $i = 1, \ldots, n$, let $\mu_i, \nu_i \in \mathcal{M}$ and $\alpha_i \geq 0$ as in Axiom 3. Without loss of generality, assume that $\sum_{i=1}^{n} \alpha_i = 1$. Define $x_{(F_i,f_i)} \geq 0$ and $y_{(F_i,F_j,f_j)} \geq 0$ by

$$x_{(F_i,f_i)} = \alpha_i \mu_i \left(p_{(F_i,f_i)}\right) \qquad \text{and} \qquad y_{(F_i,F_j,f_j)} = \alpha_i \nu_i \left(p_{(F_j,f_j)}\right) \frac{\alpha_j \mu_j \left(p_{(F_j,f_j)}\right)}{\sum_{k=1}^{n} \alpha_k \mu_k \left(p_{(F_j,f_j)}\right)}$$

where we adopt the convention that $\frac{0}{0}=0$. To follow more easily the proof, the reader may want to keep in mind the simpler case in which the supports of the μ_i are disjoint, so that $y_{(F_i,F_j,f_j)}=\alpha_i\nu_i(p_{(F_j,f_j)})$.

Simple algebra shows that

$$\sum_{i=1}^{n}\sum_{f_i\in F_i}x_{(F_i,f_i)}=\sum_{i=1}^{n}\alpha_i\sum_{f_i\in F_i}\mu_i\left(p_{(F_i,f_i)}\right)=\sum_{i=1}^{n}\alpha_i=1,$$

$$\sum_{f_i\in F_i}x_{(F_i,f_i)}p_{(F_i,f_i)}=\alpha_i\bar{\mu}_i=\alpha_i\bar{\nu}_i=\sum_{j=1}^{n}\sum_{f_j\in F_j}y_{(F_i,F_j,f_j)}p_{(F_j,f_j)},$$

$$x_{(F_j,f_j)}=\alpha_j\mu_j\left(p_{(F_j,f_j)}\right)=\alpha_j\mu_j\left(p_{(F_j,f_j)}\right)\frac{\sum_{i=1}^{n}\alpha_i\nu_i\left(p_{(F_j,f_j)}\right)}{\sum_{k=1}^{n}\alpha_k\mu_k\left(p_{(F_j,f_j)}\right)}=\sum_{i=1}^{n}y_{(F_i,F_j,f_j)}.$$

Since the linear program has non-negative value,

$$\sum_{i=1}^{n}\sum_{f_i\in F_i}x_{(F_i,f_i)}\phi_{F_i}\left(p_{(F_i,f_i)}\right)\geq\sum_{i,j=1}^{n}\sum_{f_j\in F_j}y_{(F_i,F_j,f_j)}\phi_{F_i}\left(p_{(F_j,f_j)}\right).$$

It is easy to verify that

$$\sum_{i=1}^{n}\sum_{f_i\in F_i}x_{(F_i,f_i)}\phi_{F_i}\left(p_{(F_i,f_i)}\right)=\sum_{i=1}^{n}\alpha_i\int\phi_{F_i}\,\mathrm{d}\mu_i,$$
$$\sum_{i,j=1}^{n}\sum_{f_j\in F_j}y_{(F_i,F_j,f_j)}\phi_{F_i}\left(p_{(F_j,f_j)}\right)=\sum_{i=1}^{n}\alpha_i\int\phi_{F_i}\,\mathrm{d}\nu_i.$$

We conclude that

$$\sum_{i=1}^{n}\alpha_i\int\phi_{F_i}\,\mathrm{d}\mu_i\geq\sum_{i=1}^{n}\alpha_i\int\phi_{F_i}\,\mathrm{d}\nu_i,$$

which implies that Axiom 3 is satisfied.

"Only if." Suppose Axiom 3 is satisfied. For $i,j=1,\ldots,n$, $f_i\in F_i$, and $f_j\in F_j$, take $x_{(F_i,f_i)}$ and $y_{(F_i,F_j,f_j)}$ that satisfy the constraints of the linear program. Define $\alpha_i\geq 0$ by

$$\alpha_i=\sum_{f_i\in F_i}x_{(F_i,f_i)}. \tag{47}$$

Constraint (45) implies that

$$\sum_{f_i \in F_i} x_{(F_i,f_i)} = \sum_{j=1}^{n} \sum_{f_j \in F_j} y_{(F_i,F_j,f_j)}.$$

Thus we also have

$$\alpha_i = \sum_{j=1}^{n} \sum_{f_j \in F_j} y_{(F_i,F_j,f_j)}. \tag{48}$$

Define $\mu_i \in \mathcal{M}$ and $\nu_i \in \mathcal{M}$ by

$$\mu_i\left(p_{(F_i,f_i)}\right) = \frac{x_{(F_i,f_i)}}{\alpha_i} \qquad \text{and} \qquad \nu_i(p) = \sum_{j=1}^{n} \sum_{f_j \in F_j} \frac{y_{(F_i,F_j,f_j)}}{\alpha_i} 1_{\left\{p_{(F_j,f_j)}\right\}}(p).$$

By (47) and (48), μ_i and ν_i are well defined. If $\alpha_i = 0$, we adopt the convention that $\mu_i = \nu_i = \mu_{F_i}$. To follow more easily the proof, the reader may want to keep in mind the simpler case in which the supports of the μ_{F_i} are disjoint, so that $\nu_i(p_{(F_j,f_j)}) = y_{(F_i,F_j,f_j)}/\alpha_i$.

By construction, $\mu_i(\mathcal{P}_{F_i}) = 1$. In addition, constraint (45) implies that $\bar{\mu}_i = \bar{\nu}_i$. Constraint (46) implies that $\sum_i \alpha_i \mu_i = \sum_i \alpha_i \nu_i$. Since Axiom 3 is satisfied, we obtain that

$$\sum_{i=1}^{n} \alpha_i \int \phi_{F_i} \, \mathrm{d}\mu_i \geq \sum_{i=1}^{n} \alpha_i \int \phi_{F_i} \, \mathrm{d}\nu_i.$$

It is clear that

$$\sum_{i=1}^{n} \sum_{f_i \in F_i} x_{(F_i,f_i)} \phi_{F_i}\left(p_{(F_i,f_i)}\right) = \sum_{i=1}^{n} \alpha_i \int \phi_{F_i} \, \mathrm{d}\mu_i,$$

$$\sum_{i,j=1}^{n} \sum_{f_j \in F_j} y_{(F_i,F_j,f_j)} \phi_{F_i}\left(p_{(F_j,f_j)}\right) = \sum_{i=1}^{n} \alpha_i \int \phi_{F_i} \, \mathrm{d}\nu_i.$$

We deduce that

$$\sum_{i=1}^{n} \sum_{f_i \in F_i} x_{(F_i,f_i)} \phi_{F_i}\left(p_{(F_i,f_i)}\right) \geq \sum_{i,j=1}^{n} \sum_{f_j \in F_j} y_{(F_i,F_j,f_j)} \phi_{F_i}\left(p_{(F_j,f_j)}\right).$$

This proves that the linear program has non-negative value. $\square$

Axiom 3 is an all-or-nothing condition, a common feature of Afriat-style axioms. We may be interested also in the *degree* to which the axiom is violated. We can use the equivalence between Axiom 3 and a linear program to provide a tractable measure of the severity of a violation.

To simplify notation, assume that for all $F \in \mathcal{F}$, $\operatorname{supp}(\sigma_F) = F$. For every $F \in \mathcal{F}$ and $f \in F$, let $\epsilon_{(F,f)} \in [0,1]$. Setting

$$\boldsymbol{\epsilon} = \left(\epsilon_{(F,f)} : F \in \mathcal{F}, f \in F\right),$$

we say that Axiom 3 is $\boldsymbol{\epsilon}$*-satisfied* if the following linear program has non-negative value:

$$\begin{aligned}
\text{minimize} \quad & \sum_{F\in\mathcal{F}}\sum_{f\in F} x_{(F,f)}\phi_F\left(p_{(F,f)}\right) - \sum_{F,G\in\mathcal{F}}\sum_{g\in G} y_{(F,G,g)}\phi_F\left(p_{(G,g)}\right) \\
\text{subject to} \quad & x_{(F,f)} \geq 0 \quad \text{and} \quad y_{(F,G,g)} \geq 0, \qquad F,G\in\mathcal{F}, f\in F, g\in G, \\
& \sum_{F\in\mathcal{F}}\sum_{f\in F} x_{(F,f)} = 1, \\
& \sum_{f\in F} x_{(F,f)}p_{(F,f)} = \sum_{G\in\mathcal{F}}\sum_{g\in G} y_{(F,G,g)}p_{(G,g)}, \qquad F\in\mathcal{F}, \\
& x_{(G,g)} = \sum_{F\in\mathcal{F}} y_{(F,G,g)}, \qquad G\in\mathcal{F}, g\in G, && (49) \\
& y_{(F,F,f)} \geq x_{(F,f)} - \epsilon_{(F,f)}, \qquad F\in\mathcal{F}, f\in F. && (50)
\end{aligned}$$

If $\boldsymbol{\epsilon} = \mathbf{1}$–that is, if $\epsilon_{(F,f)} = 1$ for all $F \in \mathcal{F}$ and $f \in F$–then (50) is redundant: Axiom 3 is $\mathbf{1}$-satisfied if and only if it is satisfied in the usual sense (see Proposition 5). If $\boldsymbol{\epsilon} = \mathbf{0}$–that is, if $\epsilon_{(F,f)} = 0$ for all $F \in \mathcal{F}$ and $f \in F$–then (49) and (50) imply that $y_{(F,F,f)} = x_{(F,f)}$ for all $F \in \mathcal{F}$ and $f \in F$, which in turn implies that the value of the linear program is zero: Axiom 3 is always $\mathbf{0}$-satisfied. Intuitively, as $\boldsymbol{\epsilon}$ ranges from $\mathbf{0}$ and $\mathbf{1}$, more and more reallocations of revealed posteriors are tested. If Axiom 3 is $\boldsymbol{\epsilon}$-satisfied for $\boldsymbol{\epsilon}$ close to $\mathbf{1}$ but not exactly $\mathbf{1}$, then the violation of the axiom is not too severe.

To aggregate the different dimensions of the vector $\boldsymbol{\epsilon}$, we can build a consistency index in the spirit of Afriat and Varian (see, e.g., Chambers and Echenique, 2016, ch. 5). For every $F \in \mathcal{F}$ and $f \in F$, let $\alpha_{(F,f)} \in (0,1)$ such that

$$\sum_{F\in\mathcal{F}, f\in F} \alpha_{(F,f)} = 1.$$

Given $\boldsymbol{\alpha} = \left(\alpha_{(F,f)} : F \in \mathcal{F}, f \in F\right)$, we can use the quantity

$$\max\left\{\sum_{F\in\mathcal{F}, f\in F} \alpha_{(F,f)}\epsilon_{(F,f)} : \text{Axiom 3 is } \boldsymbol{\epsilon}\text{-satisfied}\right\}$$

as a consistency index for Axiom 3. The vector of weights $\boldsymbol{\alpha}$ aggregates the the different dimensions of the vector $\boldsymbol{\epsilon}$. Computing the index boils down to a linear program.

A limitation of the consistency index is that it does not distinguish between violations of Axioms 2 and 3. It could be interested to try to separate the two types of violations, in the spirit of Halevy, Persitz and Zrill (2018).

D Uniform posterior separability

In this section, we extend the analysis to cost functions that are *uniformly* posterior separable. So far, the agent's prior has been fixed. Uniform posterior separability, due to Caplin, Dean and Leahy (2017), is an hypothesis on the cost of information *across priors*.

Let $\mathcal{M}_+$ be the set of distributions over posteriors whose barycenters have full support:

$$\mathcal{M}_+ = \{\mu \in \mathcal{M} : \bar{\mu}(\theta) > 0 \text{ for all } \theta \in \Theta\}.$$

Definition 8. A cost function $C : \mathcal{M}_+ \to (-\infty, \infty]$ is *uniformly posterior separable* if there is an upper semicontinuous function $H : \Delta(\Theta) \to \mathbb{R}$ such that for all $\mu \in \mathcal{M}_+$,

$$C(\mu) = H(\bar{\mu}) - \int_{\Delta(\Theta)} H(p)\, \mathrm{d}\mu(p).$$

A cost function is uniformly posterior separable if it is posterior separable and the measure of uncertainty is the same across priors. Uniform posterior separability is a strengthening of posterior separability that is often assumed in rational inattention; for example, it is often assumed that H is entropy regardless of the prior beliefs of the decision maker. Uniform posterior-separable costs have shown some limitations–e.g., in settings where the prior is endogenous (Ravid, 2020; Denti, Marinacci and Rustichini, 2021)–hereby motivating the need for tests.

If the analyst observes the behavior of the agent only for a fixed prior, then uniform posterior separability is observationally equivalent to posterior separability. To characterize the additional implications of uniform posterior separability, we allow the analyst to observe the agent's behavior across menus *and* across priors. Each decision problem is now identified by a finite menu of acts $F \subseteq \mathbb{R}^\Theta$ *and* by the agent's (full support) prior $\pi \in \Delta(\Theta)$. The prior may vary across decision problems.

Let $\mathbb{D}$ be a finite collection of decision problems for which the analyst observes the behavior of the agent. For every decision problem $(\pi, F) \in \mathbb{D}$, a stochastic kernel $\sigma_{(\pi,F)} : \Theta \to \Delta(F)$ describes the probability $\sigma_{(\pi,F)}(f|\theta) \in [0,1]$ that the agent selects act f from menu F in state θ. The analyst's dataset consists of a rule that associates to every decision problem $(\pi, F) \in \mathbb{D}$ a stochastic kernel $\sigma_{(\pi,F)} \in \mathcal{S}_F$.

Let $\sigma_{(\mu,d_F)} \in \mathcal{S}_F$ be the stochastic kernel induced by a pair $(\mu, d_F) \in \mathcal{M}_+ \times \mathcal{D}_F$:

$$\sigma_{(\mu,d_F)}(f|\theta) = \int_{\Delta(\Theta)} \frac{d_F(f|p)p(\theta)}{\bar{\mu}(\theta)} \, \mathrm{d}\mu(p).$$

Definition 9. A cost function $C : \mathcal{M}_+ \to (-\infty, \infty]$ *rationalizes* the dataset $\big(\sigma_{(\pi,F)} : (\pi, F) \in \mathbb{D}\big)$ if for every $(\pi, F) \in \mathbb{D}$, there is a solution $(\mu, d_F) \in \mathcal{M}_\pi \times \mathcal{D}_F$ of (1) such that $\sigma_{(\pi,F)} = \sigma_{(\mu,d_F)}$.

An extension of our main axioms characterizes the cost functions that are uniformly posterior separable. To illustrate, let $p_{(\pi,F,f)}$ be the revealed posterior for an act $f \in F$ in a decision problem $(\pi, F) \in \mathbb{D}$. Denote by $\mathcal{P}_{(\pi,F)}$ the set of revealed posteriors and by $\mu_{(\pi,F)}$ the distribution of revealed posteriors.

Axiom 4. *For every $(\pi, F) \in \mathbb{D}$, $f \in \operatorname{supp}(\sigma_{(\pi,F)})$, and $g \in F$,*

$$f \cdot p_{(\pi,F,f)} \geq g \cdot p_{(\pi,F,f)} \qquad \forall g \in F.$$

Axiom 5. *For every $i = 1, \ldots, n$, let $(\pi_i, F_i) \in \mathbb{D}$, $\alpha_i \in \mathbb{R}_+$, and $\mu_i \in \mathcal{M}$ such that $\mu_i(\mathcal{P}_{(\pi_i,F_i)}) = 1$. For every reallocation of posteriors $\nu_1, \ldots, \nu_n \in \mathcal{M}$,*

$$\sum_{i=1}^{n} \alpha_i \int \phi_{F_i} \, \mathrm{d}\mu_i \geq \sum_{i=1}^{n} \alpha_i \int \phi_{F_i} \, \mathrm{d}\nu_i.$$

Axiom 4 is the same as Axiom 1. Axiom 5, instead, is a substantial strengthening of Axiom 3: by Axiom 5, the agent's total utility cannot be improved by reallocating revealed posteriors across menus *and* across priors. Axioms 4 and 5 characterize the revealed preference implications of uniform posterior separability:

Theorem 4. *A dataset $(\sigma_{(\pi,F)} : (\pi, F) \in \mathbb{D})$ satisfies Axioms 4 and 5 if and only if it is rationalized by a cost function that is uniformly posterior separable. In addition, we can choose H to be concave.*

The proof (omitted) follows the exact same steps of the proof of Theorem 2. With suitable data, experimental tests of Axiom 4 can be conducted in the same fashion as the tests for Axiom 3 we describe in Section 4. Next we provide an example:

Example 2. Modify the setting of Section 4.2as follows: as the index i increases, not only the stakes increase, but also the prior becomes more dogmatic:

$$1/2 \leq \pi_1 < \pi_2 < \pi_3 < \pi_4 < 1.$$

As i increases, it becomes ex-ante more likely that there are more red dots on the screen.

As the index i increases, the change in the incentive to acquire information is *ambiguous*. On one hand, the stakes increase; this suggests that the incentive to acquire information is *higher* for higher indices. At the same time, as the index i increases, the prior becomes more dogmatic; this suggests that the incentive to acquire information is *lower* for higher indices. It is up to the decision maker to aggregate these countervailing factors.

It is easy to modify the proof of Proposition 1 to obtain the following result: Given a pair of decision problems i and j with $i < j$ and $\pi_i = 1/2 < \pi_j$, Axiom 5 is satisfied if and only if for all $p \in \mathcal{P}_{(\pi_j, F_j)}$,

$$\left(p_{(\pi_i,F_i,f_i)} - p\right)\left(p - p_{(\pi_i,F_i,g_i)}\right) \leq 0. \tag{51}$$

Thus, Axiom 5 is satisfied if and only if as the index i increases, the revealed posterior beliefs become more extreme.

Condition (51) can be quite demanding. Fox example, take $i = 3$ and $j = 4$, so that the stakes increase by a relatively small amount. Imagine that given decision problem (π_3, F_3), the agent finds it optimal to acquire full information, that is, $p_{(\pi_3,F_3,f_3)} = 1$ and $p_{(\pi_3,F_3,g_3)} = 0$. Then, no matter how close π_4 is to one, (51) requires that $p_{(\pi_4,F_4,f_4)} = 1$ and $p_{(\pi_4,F_4,g_4)} = 0$: Given decision problem (π_4, F_4), the agent *must* acquire full information. This seems counterintuitive since almost dogmatic prior beliefs could dissuade the agent to acquire any information at all.

Condition (51) shows that, in this context, testing Axiom 5 is equivalent to testing a system of moment inequalities. Thus, with suitable data, one could test Axiom 5 using the same method of inference we use in the empirical application of Section 4. The experimental design of Dean and Neligh (2019) is flexible enough to accommodate heterogenous priors.[26]

E Rich datasets: identification and comparative statics

In this section we drop the hypothesis that $\mathcal{F}$ is finite to study identification and comparative statics in rich datasets. Among the results in the main text, Theorems 1 and 3 easily extend to the case where $\mathcal{F}$ is infinite: the extension of Theorem 1 to infinite datasets is discussed by Caplin, Dean and Leahy (2017); The proof of Theorem 3 extends verbatim to the case where $\mathcal{F}$ is infinite.

The extension of Theorem 2 to infinite datasets is less straightforward. The proof of Theorem 2 invokes a finite-dimensional version of the theorem of the alternative (i.e., Farkas'

[26] One of the experiments run by Dean and Neligh (2019) feature heterogenous priors. Such experiment, however, cannot be used to test Axiom 5 because the menu is constant across decision problems. If $F_1 = \ldots = F_n$, then Axiom 5 is trivially satisfied, even if $\pi_1 \neq \ldots \neq \pi_n$.

lemma). For infinite datasets, one would need to consider either an infinite-dimensional version of the theorem of the alternative, which usually comes with additional technical conditions, or a different proof strategy.

E.1 Test functions

Test functions, in the spirit of Lu (2016), are the tool we need to study identification and comparative statics in rich datasets. For $F \in \mathcal{F}$, act $g \in \mathbb{R}^\Theta$, and $\alpha \in (0,1)$, define the menu $\alpha F + (1-\alpha)g$ by

$$\alpha F + (1-\alpha)g = \{\alpha f + (1-\alpha)g : f \in F\}.$$

Definition 10. Let $F \in \mathcal{F}$ and $g \in \mathbb{R}^\Theta$ such that for every $\alpha \in (0,1)$, $\alpha F + (1-\alpha)g \in \mathcal{F}$. The (F,g)-*test function* is the function $T_{(F,g)} : (0,1) \to \mathbb{R}_+$ given by

$$T_{(F,g)}(\alpha) = \frac{1}{\alpha}\left(\int_{\Delta(\Theta)} \phi_{(\alpha F+(1-\alpha)g)}(p)\,\mathrm{d}\mu_{(\alpha F+(1-\alpha)g)}(p) - \phi_{(\alpha F+(1-\alpha)g)}(\pi)\right).$$

Test functions address the following question: How does the value of the agent's information changes as the incentive to acquire information changes? The term

$$\int_{\Delta(\Theta)} \phi_{(\alpha F+(1-\alpha)g)}(p)\,\mathrm{d}\mu_{(\alpha F+(1-\alpha)g)}(p) - \phi_{(\alpha F+(1-\alpha)g)}(\pi)$$

is the value of the revealed information structure $\mu_{(\alpha F+(1-\alpha)g)}$ for the menu $\alpha F + (1-\alpha)g$. Changing α means changing the incentive to acquire information. Intuitively, as α decreases toward zero, the mixture menu $\alpha F + (1-\alpha)g$ becomes more similar to the singleton menu $\{g\}$, for which information is useless.

The next lemma describes the main properties of test functions. To state the result, given a cost function C and a menu F, we denote by $V_C(F)$ the value that, ex ante, the agent assigns to menu F:

$$V_C(F) = \max_{\mu \in \mathcal{M}_\pi} \int_{\Delta(\Theta)} \phi_F(p)\,\mathrm{d}\mu(p) - C(\mu).$$

Lemma 11. *Suppose that the dataset $(\sigma_F)_{F\in\mathcal{F}}$ is rationalized by a canonical cost function C. Then every test function $T_{(F,g)}$ is non-decreasing and satisfies*

$$V_C(F) = \phi_F(\pi) + \int_0^1 T_{(F,g)}(\alpha)\,\mathrm{d}\alpha.$$

In addition, every two test functions $T_{(F,g)}$ and $T_{(F,g')}$ are equal almost everywhere.

The lemma highlights three important properties of test functions. First, test functions are non-decreasing: if $\alpha \geq \beta$ then $T_{(F,g)}(\alpha) \geq T_{(F,g)}(\beta)$. For the ratio

$$T_{(F,g)}(\alpha) = \frac{\int_{\Delta(\Theta)} \phi_{(\alpha F+(1-\alpha)g)}(p)\,\mathrm{d}\mu_{(\alpha F+(1-\alpha)g)}(p) - \phi_{(\alpha F+(1-\alpha)g)}(\pi)}{\alpha}$$

to be non-decreasing in α, the numerator must increase faster than the denominator. This is evidence that the agent is responsive to the incentive to acquire information.

Second, test functions allow us to recover the ex ante preferences of the agent over menus. Imagine that, before acquiring any information, the agent is given the opportunity to choose between the menus F and G. The agent would prefer F to G if and only if $V_C(F) \geq V_C(G)$. These preferences over menus are studied by de Oliveira, Denti, Mihm and Ozbek (2017); we will build on their identification result.

Third, fixing a menu F, every two test functions $T_{(F,g)}$ and $T_{(F,g')}$ are equal almost everywhere. This reflects the fact that the menus $\alpha F+(1-\alpha)g$ and $\alpha F+(1-\alpha)g'$ generate the same incentive to acquire information. It would not be true if the utility function and the cost of information were not additively separable, as in the model of Chambers, Liu and Rehbeck (2020).

Proof. Let $\mathcal{C}(\Delta(\Theta))$ be the space of continuous functions $\phi : \Delta(\Theta) \to \mathbb{R}$. Define the functional $W_C : \mathcal{C}(\Delta(\Theta)) \to \mathbb{R}$ by

$$W_C(\phi) = \max_{\mu \in \mathcal{M}_\pi} \int \phi \,\mathrm{d}\mu - C(\mu).$$

The functional W_C is convex and continuous in the sup-norm topology. Moreover, for every menu F, $W_C(\phi_F) = V_C(F)$.

Let $F \in \mathcal{F}$ and $g \in \mathbb{R}^\Theta$ such that for every $\alpha \in (0,1)$, $\alpha F + (1-\alpha)g \in \mathcal{F}$. Define the function $\Psi_{(F,g)} : \mathbb{R} \to \mathbb{R}$ by

$$\Psi_{(F,g)}(\alpha) = W_C\left(\alpha\phi_F + (1-\alpha)\phi_{\{g\}}\right).$$

Since W_C is convex, the function $\Psi_{(F,g)}$ is convex.

Let $\Psi^+_{(F,g)}$ and $\Psi^-_{(F,g)}$ be the right and left derivatives of $\Psi_{(F,g)}$. Observe that for all $\alpha \in (0,1)$,

$$\alpha\phi_F + (1-\alpha)\phi_{\{g\}} = \phi_{(\alpha F+(1-\alpha)g)}.$$

Since C is monotone and rationalizes the dataset, it follows from Lemma 3 that for all $G \in \mathcal{F}$,

$$\mu_G \in \arg\max_{\mu \in \mathcal{M}_\pi} \int \phi_G \,\mathrm{d}\mu - C(\mu).$$

Thus, by an envelope theorem (e.g, Milgrom and Segal 2002, Theorem 1), we have that for all $\alpha \in (0,1)$,

$$\Psi^{-}_{(F,g)}(\alpha) \leq \int \phi_F - \phi_{\{g\}} \, \mathrm{d}\mu_{(\alpha F + (1-\alpha) g)} \leq \Psi^{+}_{(F,g)}(\alpha).$$

Simple algebra shows that

$$\int \phi_F - \phi_{\{g\}} \, \mathrm{d}\mu_{(\alpha F + (1-\alpha) g)} = T_{(F,g)}(\alpha) + \phi_F(\pi) - \phi_{\{g\}}(\pi).$$

We deduce that

$$\Psi^{-}_{(F,g)}(\alpha) \leq T_{(F,g)}(\alpha) + \phi_F(\pi) - \phi_{\{g\}}(\pi) \leq \Psi^{+}_{(F,g)}(\alpha). \tag{52}$$

By Rockafellar (1970, Theorem 24.1), if $\alpha < \beta$ then $\Psi^{+}_{(F,g)}(\alpha) \leq \Psi^{-}_{(F,g)}(\beta)$. Putting this together with (52), we obtain that $T_{(F,g)}$ is increasing. In addition, Rockafellar (1970, Theorem 25.3) guarantees that $\Psi^{+}_{(F,g)} = \Psi^{-}_{(F,g)}$ almost everywhere. It follows from (52) that for almost all $\alpha \in (0,1)$,

$$\Psi^{+}_{(F,g)}(\alpha) = T_{(F,g)}(\alpha) + \phi_F(\pi) - \phi_{\{g\}}(\pi). \tag{53}$$

By Rockafellar (1970, Corollary 24.2.1) we have that

$$V_C(F) - \phi_{\{g\}}(\pi) = \Psi_{(F,g)}(1) - \Psi_{(F,g)}(0) = \int_0^1 \Psi^{+}_{(F,g)}(\alpha) \, \mathrm{d}\alpha.$$

We conclude that

$$V_C(F) = \phi_F(\pi) + \int_0^1 T_{(F,g)}(\alpha) \, \mathrm{d}\alpha.$$

Let $g' \in \mathbb{R}^{\Theta}$ such that for every $\alpha \in (0,1)$, $\alpha F + (1-\alpha) g' \in \mathcal{F}$. Note that for all $\alpha \in \mathbb{R}$,

$$\Psi_{(F,g)}(\alpha) - (1-\alpha)\phi_{\{g\}}(\pi) = \Psi_{(F,g')}(\alpha) - (1-\alpha)\phi_{\{g'\}}(\pi).$$

This implies that

$$\Psi^{+}_{(F,g)}(\alpha) + \phi_{\{g\}}(\pi) = \Psi^{+}_{(F,g')}(\alpha) + \phi_{\{g'\}}(\pi).$$

We conclude from (53) that $T_{(F,g)} = T_{(F,g')}$ almost everywhere. $\square$

E.2 Rich datasets

Test functions allows us to uniquely identify canonical costs when the dataset is "rich." Next we formalize a notion of richness. Let $\mathrm{co}(F)$ be the *convex hull* of menu $F = \{f_1, \ldots, f_n\}$:

$$\mathrm{co}(F) = \left\{ \sum_{i=1}^{n} \alpha_i f_i : \alpha_1, \ldots, \alpha_n \geq 0 \text{ and } \sum_{i=1}^{n} \alpha_i = 1 \right\}.$$

Let $d(\mathrm{co}(F), \mathrm{co}(G))$ be the *Hausdorff distance* of the convex hulls of menus F and G:

$$d(\mathrm{co}(F), \mathrm{co}(G)) = \max\left\{ \max_{f \in \mathrm{co}(F)} \min_{g \in \mathrm{co}(G)} \|f - g\|, \max_{g \in \mathrm{co}(G)} \min_{f \in \mathrm{co}(F)} \|g - f\| \right\}.$$

Definition 11. The collection of menus $\mathcal{F}$ is *rich* if the following conditions hold:

(i) For every $\epsilon > 0$ and $G \subseteq \mathbb{R}^{\Theta}$, there is $F \in \mathcal{F}$ such that

$$d(co(F), co(G)) \leq \epsilon.$$

(ii) For every $F \in \mathcal{F}$, there is $g \in \mathbb{R}^{\Theta}$ such that for every $\alpha \in (0, 1)$,

$$\alpha F + (1 - \alpha) g \in \mathcal{F}.$$

By (i), the collection of menus in the dataset is dense in the collection of all menus. The distance between menus is the Hausdorff distance between the convex hulls. The convex hull of a menu represents the set of feasible payoffs, provided that mixing is allowed.

Condition (ii) guarantees that for every menu F, we are able to construct a test function $T_{(F,g)}$ for some act g. The identity of g is irrelevant since every two test functions $T_{(F,g)}$ and $T_{(F,g')}$ are equal almost everywhere (see Lemma 11). For example, g could simply be an element of F. An example of a rich dataset is, of course, the collection of all menus.

E.3 Unique identification

The next theorem shows that a rich dataset can be rationalized by one and only one canonical cost function.

Theorem 5. *Suppose that the dataset $(\sigma_F)_{F \in \mathcal{F}}$ is rationalized by a cost function C. If $\mathcal{F}$ is rich, then the dataset is rationalized by a unique canonical cost function C^* given by*

$$C^*(\mu) = \sup\left\{ \int \phi_{F_n}\, \mathrm{d}\mu + \sum_{i=1}^{n-1} \int \phi_{F_i}\, \mathrm{d}\mu_{F_{i+1}} - \sum_{i=1}^{n} \int \phi_{F_i}\, \mathrm{d}\mu_{F_i} \right\}$$

where the supremum is taken over all finite sequences $F_1, \ldots, F_n$ in $\mathcal{F}_{\{0\}}$.

If the dataset is rich, then there is a *unique canonical* cost function that rationalizes the dataset. We emphasize that identification is achieved by focusing on canonical costs. There could be other cost functions that are not canonical, that rationalize the dataset, and that are different from C^*. As for Theorem 3, the hypothesis that the cost is canonical can be seen as a normalization to achieve identification.

The proof of Theorem 5 builds on the identification result for menu choices obtained by de Oliveira, Denti, Mihm and Ozbek (2017). The next lemma immediately follows from their work.

Lemma 12. *If C is canonical and $\mathcal{F}$ is rich, then*

$$C(\mu) = \sup_{F \in \mathcal{F}} \int \phi_F \, \mathrm{d}\mu - V_C(F).$$

Proof. By de Oliveira, Denti, Mihm and Ozbek (2017, Theorem 2),

$$C(\mu) = \sup \left\{ \int \phi_F \, \mathrm{d}\mu - V_C(F) : F \text{ is a finite subset of } \mathbb{R}^\Theta \right\}.$$

Since $\mathcal{F}$ is rich, for every finite subset F of $\mathbb{R}^\Theta$ and $\epsilon > 0$, there is F_ϵ such that

$$d\left(\mathrm{co}(F), \mathrm{co}(F_\epsilon)\right) \leq \epsilon.$$

By Schneider (2014, Lemma 1.8.14),

$$\max_{p \in \Delta(\Theta)} |\phi_F(p) - \phi_{F_\epsilon}(p)| \leq \epsilon.$$

The desired result follows from the continuity of V_C. □

Proof of Theorem 5. In view of Theorem 3, it is enough to show that the dataset is represented by a unique canonical cost function. Let C_1 and C_2 be two canonical cost functions representing the dataset. By Lemma 11, we have that for every $F \in \mathcal{F}$,

$$V_{C_1}(F) = \phi_F(\pi) + \int_0^1 T_{(F,g)}(\alpha) \, \mathrm{d}\alpha = V_{C_2}(F).$$

Thus, by Lemma 12, we obtain that for every $\mu \in \mathcal{M}_\pi$,

$$C_1(\mu) = \sup_{F \in \mathcal{F}} \int \phi_F \, \mathrm{d}\mu - V_{C_1}(F) = \sup_{F \in \mathcal{F}} \int \phi_F \, \mathrm{d}\mu - V_{C_2}(F) = C_2(\mu).$$

□

Applied to posterior separable costs, Theorem 5 allows us to identify the measure of uncertainty, up to affine translations.

Corollary 2. *Let $(\sigma_F)_{F\in\mathcal{F}}$ be a dataset of stochastic choices, with $\mathcal{F}$ rich. Let C_1 and C_2 be two posterior separable costs, with concave information measures H_1 and H_2. If C_1 and C_2 both rationalize the dataset, then there is $\zeta \in \mathbb{R}^\Theta$ such that for all $p \in \Delta(\Theta)$,*

$$H_1(p) = H_2(p) + \zeta \cdot p.$$

To prove the result, one only need to combine Lemma 2 and Theorem 5.

E.4 Comparative statics

Test functions also allow us to characterize the behavioral implications of increasing the cost of information. We compare two agents with costs given by C_1 and C_2, respectively. We assume that both C_1 and C_2 are canonical.

Definition 12. Agent 1 has a higher cost of information than agent 2 if $C_1 \geq C_2$.

If we specialize the definition to posterior separable costs, we obtain a statement about the relative concavity of the measures of uncertainty:

Definition 13. A function $\psi : \Delta(\Theta) \to \mathbb{R}$ *satisfies Jensen's inequality* at π if for all finite sequences $p_1, \ldots, p_n$ in $\Delta(\Theta)$ and $\alpha_1, \ldots, \alpha_n$ in $[0, 1]$ such that $\sum_{i=1}^n \alpha_i p_i = \pi$,

$$\sum_{i=1}^{n} \alpha_i \psi(p_i) \geq \psi(\pi).$$

Lemma 13. *Assume that C_1 and C_2 are posterior separable with concave uncertainty measures H_1 and H_2. Agent 1 has higher cost of information than agent 2 if and only if the function $\psi = H_2 - H_1$ satisfies Jensen's inequality at π.*

Proof. "If." Assume that ψ satisfies Jensen's inequality at the prior. If μ has finite support, then

$$C_1(\mu) - C_2(\mu) = \sum_p \mu(p)\psi(p) - \psi(\pi) \geq 0.$$

□

By Winkler (1988, Theorem 2.1), the extreme points of $\mathcal{M}_\pi$ have finite support. Thus, by the Krein-Milman theorem (Aliprantis and Border, 2006, Theorem 7.68), every $\mu \in \mathcal{M}_\pi$ is the limit of a sequence (μ_n) in $\mathcal{M}_\pi$ such that each μ_n has finite support. We conclude that, by continuity of the cost functions, the inequality $C_2(\mu) \geq C_1(\mu)$ extends to all $\mu \in \mathcal{M}_\pi$.

"Only if." Assume that $C_2 \geq C_1$. Let $p_1, \ldots, p_n$ in $\Delta(\Theta)$ and $\alpha_1, \ldots, \alpha_n$ in $[0,1]$ such that $\sum_{i=1}^{n} \alpha_i p_i = \pi$. Define $\mu \in \mathcal{M}_\pi$ such that $\mu = \sum_i \alpha \delta_{p_i}$. Then

$$\sum_i \alpha_i \psi(p_i) - \psi(\pi) = C_1(\mu) - C_2(\mu) \geq 0.$$

This shows that ψ satisfies Jensen's inequality at the prior.

A concrete example of the comparative statics we study here often appears in applications of rational inattention.

Example 3. Let H be a measure of uncertainty. Given $\alpha_1, \alpha_2 \geq 0$, define $H_1 = \alpha_1 H$ and $H_2 = \alpha_2 H$. Suppose that agents 1 and 2 have posterior separable costs with uncertainty measures H_1 and H_2. Then agent 1 has higher cost of information than agent 2 if and only if $\alpha_1 \geq \alpha_2$.

Next we use test functions to provide a behavioral characterization of increasing the cost of information. Let $T^1_{(F,g)}$ and $T^2_{(F,g)}$ be the (F,g)-test functions of agents 1 and 2.

Definition 14. The test function $T^1_{(F,g)}$ *second-order stochastically dominates* $T^2_{(F,g)}$ if for all $\alpha \in (0,1]$,

$$\int_0^\alpha T^2_{(F,g)}(\beta)\,\mathrm{d}\beta \geq \int_0^\alpha T^1_{(F,g)}(\beta)\,\mathrm{d}\beta.$$

Theorem 6. *Suppose that $\mathcal{F}$ is rich. The following statements are equivalent:*

(i) Agent 1 has higher cost of information than agent 2.

(ii) For every $F \in \mathcal{F}$, $T^1_{(F,g)}$ second-order stochastically dominates $T^2_{(F,g)}$.

Thus, increasing the cost of information means increasing the test function, in the sense of second-order stochastic dominance. From Lemma 11, test functions are non-decreasing non-negative functions from the interval $(0,1)$ into the real line. Test functions, therefore, can be seen as cumulative distribution functions and ranked by second-order stochastic dominance. Intuitively, an *increase* in second-order stochastic dominance for test functions reveals a *decrease* in sensitivity to the incentive to acquire information, hence a higher cost of information.

Proof. "(i) implies (ii)." Conditions (i) implies that for all $F \in \mathcal{F}$,

$$V_{C_2}(F) \geq V_{C_1}(F).$$

By Lemma 11, we have that for all $F \in \mathcal{F}$,

$$\int_0^1 T^1_{(F,g)}(\alpha)\,\mathrm{d}\alpha \geq \int_0^1 T^2_{(F,g)}(\alpha)\,\mathrm{d}\alpha. \tag{54}$$

Observe that for all $F \in \mathcal{F}$, $\alpha \in (0,1]$, and $\beta \in (0,1)$,

$$\beta(\alpha F + (1-\alpha)g) + (1-\beta)g = \alpha\beta F + (1-\alpha\beta)g.$$

As a result, with a change of variable we get that

$$\begin{aligned}
&\int_0^1 T^i_{(\alpha F+(1-\alpha)g,g)}(\beta)\,\mathrm{d}\beta \\
=&\int_0^1 \frac{1}{\beta}\left(\int \phi_{(\alpha\beta F+(1-\alpha\beta)g)}\,\mathrm{d}\mu_{(\alpha\beta F+(1-\alpha\beta)g)} - \phi_{(\alpha\beta F+(1-\alpha\beta)g)}(\pi)\right)\mathrm{d}\beta \\
=&\frac{1}{\alpha}\int_0^\alpha \frac{1}{\beta}\left(\int \phi_{(\beta F+(1-\beta)g)}\,\mathrm{d}\mu_{(\beta F+(1-\beta)g)} - \phi_{(\beta F+(1-\beta)g)}(\pi)\right)\mathrm{d}\beta \\
=&\frac{1}{\alpha}\int_0^\alpha T^i_{(F,g)}(\beta)\,\mathrm{d}\beta. \qquad (55)
\end{aligned}$$

Then (ii) follows from (54) and (55).

"(ii) implies (i)." By Lemma 11, condition (ii) implies that for all $F \in \mathcal{F}$,

$$V_{C_2}(F) \geq V_{C_1}(F).$$

By Lemma 12, we conclude that $C_1 \geq C_2$. □

E.5 Related literature

Lu (2016) inspired us to use test functions to study identification and comparative statics. Lu takes the perspective of an analyst who observes the behavior of an agent across menus $F \subset \mathbb{R}^\Theta$. In his paper, the analyst's dataset is a rule $(\bar{\sigma}_F)_{F\in\mathcal{F}}$ that associates to each menu $F \in \mathcal{F}$ a *state-independent* stochastic choice $\bar{\sigma}_F \in \Delta(F)$. The stochastic choice $\bar{\sigma}_F$ can be seen as the average of an underlying *state-dependent* stochastic choice $\sigma_F : \Theta \to \Delta(F)$:

$$\bar{\sigma}_F(f) = \sum_\theta \sigma_F(f|\theta)\pi(\theta).$$

It should be noted that Lu does not assume that the analyst knows the agent's utility or his prior. For ease of exposition, we do no emphasize this difference between the papers.

Lu considers a choice model where the agent's private information is exogenous. The agent's information is represented by a *fixed* $\mu^* \in \mathcal{M}_\pi$. Given a menu F of feasible acts, the agent chooses $d_F \in \mathcal{D}_F$ to maximize

$$\int_{\Delta(\Theta)} \left(\sum_f (f\cdot p) d_F(f|p)\right) \mathrm{d}\mu^*(p).$$

The pair (μ^*, d_F) induces the state-independent stochastic choice $\bar{\sigma}_{(\mu^*,d_F)} \in \Delta(F)$ given by

$$\bar{\sigma}_{(\mu^*,d_F)}(f) = \sum_{\theta} \sigma_{(\mu^*,d_F)}(f|\theta)\pi(\theta).$$

The choice model of Lu is a special case of ours: take cost function C given by

$$C(\mu) = \begin{cases} 0 & \text{if } \mu^* \succeq \mu, \\ \infty & \text{otherwise.} \end{cases}$$

Lu focuses on identification and develops a notion of test function. To illustrate, suppose that $\mathcal{F}$ is the collection of all finite subsets of $[0,1]^{\Theta}$. For every $\alpha \in [0,1]$, let f_α be the constant act that takes value α in every state. To each menu $F \in \mathcal{F}$, Lu associates the test function $J_F : [0,1] \to [0,1]$ given by

$$J_F(\alpha) = \bar{\sigma}_{F \cup \{f_\alpha\}}(f_\alpha).$$

The quantity $J_F(\alpha)$ is the probability that act f_α is selected from menu $F \cup \{f_\alpha\}$. Note that, as α increases, the act f_α becomes more attractive. Intuitively, the *speed* at which f_α becomes more attractive reveals the agent's private information.

Motivated by the different setup, we have developed a different notion of test function with respect to Lu. There are, however, important analogies. In particular, in both papers test functions allow to recover the agent's ex ante preference over menus.

Lin (2019) adopts Lu's test functions to study rational inattention. Lin considers a dataset $(\bar{\sigma}_F)_{F \in \mathcal{F}}$ of state-independent stochastic choices (as Lu does), and studies a choice model with endogenous information acquisition (as we do). In a "rich" dataset, Lin exploits Lu's test functions to identify the minimal canonical cost function that rationalizes the dataset. For Lin, a dataset is "rich" if $\mathcal{F}$ consists of *all* finite subsets of a given compact convex set $K \subseteq \mathbb{R}^{\Theta}$.

Overall, Lin's and our identification results complement each other. With respect to our setting, Lin does not require the analyst to know the state of nature. Lin also does not require the analyst to know the utility function and the prior of the agent. At the same time, however, we are able to identify the *minimal canonical* cost function in *arbitrary datasets* (which can be finite), while Lin requires *rich datasets* (which are necessarily infinite). In rich datasets, we are able to identify the *unique canonical* cost function.

Caplin, Csaba, Leahy and Nov (2020) also propose a methodology to identify the cost of information, on the basis of an analogy between rational inattention and production theory; they also provide an implementation on experimental data. Similarly to Lin (2019), their analysis requires a suitably rich dataset of stochastic choices. The distinctive feature of our methodology is that it is fully non-parametric and can be implemented on arbitrary datasets. Focusing on a specific class of cost functions, Dewan and Neligh (2020) also estimate features of the cost of information from experimental data.

F Proofs omitted from the main text

Proposition 6. *In Experiment 1, given a pair of decision problems i and j with $i < j$, Axiom 2 is satisfied if and only if*

$$\left|\sigma_{F_j}(f_j|r) - \sigma_{F_j}(f_j|b)\right| \geq |\sigma_{F_i}(f_i|r) - \sigma_{F_i}(f_i|b)| .$$

Proof. Define $z_i = f_i(r)$ and $z_j = f_j(r)$. We claim that

$$\int \phi_{F_j} \,\mathrm{d}\mu_{F_i} = \frac{z_j}{2}\left(1 + |\sigma_{F_i}(f_i|r) - \sigma_{F_i}(f_i|b)|\right). \tag{56}$$

If $\mu_{F_i}(p_{(F_i,f_i)}) \in \{0,1\}$, then

$$\int \phi_{F_j} \,\mathrm{d}\mu_{F_i} = \frac{z_j}{2} \quad \text{and} \quad |\sigma_{F_i}(f_i|r) - \sigma_{F_i}(f_i|b)| = 0.$$

It follows that (56) holds. Suppose now that $\mu_{F_i}(p_{(F_i,f_i)}) \in (0,1)$. Simple algebra shows that

$$\int \phi_{F_j} \,\mathrm{d}\mu_{F_i} = \frac{z_j}{2}\left(\max\left\{\sigma_{F_i}(f_i|r), \sigma_{F_i}(f_i|b)\right\} + \max\left\{\sigma_{F_i}(g_i|b), \sigma_{F_i}(g_i|r)\right\}\right).$$

Note that $\sigma_{F_i}(f_i|r) \geq \sigma_{F_i}(f_i|b)$ if and only if $\sigma_{F_i}(g_i|b) \geq \sigma_{F_i}(g_i|r)$. Thus

$$\begin{aligned}\int \phi_{F_j} \,\mathrm{d}\mu_{F_i} &= \frac{z_j}{2} \max\left\{\sigma_{F_i}(f_i|r) + \sigma_{F_i}(g_i|b), \sigma_{F_i}(f_i|b) + \sigma_{F_i}(g_i|r)\right\} \\ &= \frac{z_j}{2} \max\left\{\sigma_{F_i}(f_i|r) + 1 - \sigma_{F_i}(g_i|r), \sigma_{F_i}(f_i|b) + 1 - \sigma_{F_i}(g_i|r)\right\}.\end{aligned}$$

It follows from simple algebra that (56) holds.

By definition, Axiom 2 holds if and only if

$$\int \phi_{F_i} \,\mathrm{d}\mu_{F_i} + \int \phi_{F_j} \,\mathrm{d}\mu_{F_j} \geq \int \phi_{F_i} \,\mathrm{d}\mu_{F_j} + \int \phi_{F_j} \,\mathrm{d}\mu_{F_i}.$$

By (56), an equivalent condition is

$$\begin{aligned}& z_i\left(|\sigma_{F_i}(f_i|r) - \sigma_{F_i}(f_i|b)|\right) + z_j\left(\left|\sigma_{F_j}(f_j|r) - \sigma_{F_j}(f_j|b)\right|\right) \\ \geq\ & z_i\left(\left|\sigma_{F_j}(f_j|r) - \sigma_{F_j}(f_j|b)\right|\right) + z_j\left(|\sigma_{F_i}(f_i|r) - \sigma_{F_i}(f_i|b)|\right),\end{aligned}$$

which in turn simplifies to

$$(z_i - z_j)\left(|\sigma_{F_i}(f_i|r) - \sigma_{F_i}(f_i|b)|\right) \geq (z_i - z_j)\left(\left|\sigma_{F_j}(f_j|r) - \sigma_{F_j}(f_j|b)\right|\right),$$

which, becuase $z_i < z_j$, is equivalent to

$$\left|\sigma_{F_j}(f_j|r) - \sigma_{F_j}(f_j|b)\right| \geq |\sigma_{F_i}(f_i|r) - \sigma_{F_i}(f_i|b)|.$$

$\square$

Proposition 7. *In each treatment of Experiment 2, Axiom 2 is satisfied if and only if*

$$\sum_{f_2} \max\left\{\sigma_{F_2}(f_2|r) - \sigma_{F_2}(f_2|b), 0\right\} \geq \sum_{f_1} \max\left\{\sigma_{F_1}(f_1|r) - \sigma_{F_1}(f_1|b), 0\right\}.$$

Proof. For all $i, j = 1, 2$, we have that

$$\begin{aligned}\int \phi_{F_j}\, \mathrm{d}\mu_{F_i} &= \sum_{p \leq 1/2} \phi_{F_j}(p)\mu_{F_i}(p) + \sum_{p > 1/2} \phi_{F_j}(p)\mu_{F_i}(p) \\ &= \sum_{p \leq 1/2} \phi_{F_1}(p)\mu_{F_i}(p) + \sum_{p > 1/2} \phi_{F_j}(p)\mu_{F_i}(p).\end{aligned}$$

Thus, Axiom 2 is satisfied if and only if

$$\sum_{p>1/2} \phi_{F_1}(p)\mu_{F_1}(p) + \sum_{p>1/2} \phi_{F_2}(p)\mu_{F_2}(p) \geq \sum_{p>1/2} \phi_{F_1}(p)\mu_{F_2}(p) + \sum_{p>1/2} \phi_{F_2}(p)\mu_{F_1}(p).$$

For all $p > 1/2$, $\phi_{F_1}(p) = 50$ and $\phi_{F_2}(p) = 100p$. Hence, the above inequality simplifies to

$$\sum_{p>1/2} \left(p - \frac{1}{2}\right)\mu_{F_2}(p) \geq \sum_{p>1/2} \left(p - \frac{1}{2}\right)\mu_{F_1}(p).$$

Notice that

$$\begin{aligned}\sum_{p>1/2} \left(p - \frac{1}{2}\right)\mu_{F_i}(p) &= \sum_{p} \max\left\{p - \frac{1}{2}, 0\right\}\mu_{F_i}(p) \\ &= \sum_{f_i} \max\left\{p_{(F_i,f_i)}\mu_{F_i}\left(p_{(F_i,f_i)}\right) - \frac{1}{2}\mu_{F_i}\left(p_{(F_i,f_i)}\right), 0\right\} \\ &= \frac{1}{4}\sum_{f_i} \max\left\{\sigma_{F_i}(f_i|r) - \sigma_{F_i}(f_i|b), 0\right\}.\end{aligned}$$

Putting everything together, we obtain that Axiom 2 is satisfied if and only if

$$\sum_{f_2} \max \{\sigma_{F_2}(f_2|r) - \sigma_{F_2}(f_2|b), 0\} \geq \sum_{f_1} \max \{\sigma_{F_1}(f_1|r) - \sigma_{F_1}(f_1|b), 0\} .$$

□

G Relation with Caplin, Dean and Leahy (2017)

In a dataset of state-dependent stochastic choices, Caplin, Dean and Leahy (2017) study the testable implications of the entropy cost cost (Matejka and McKay, 2015). Among other results, they propose characterizations of posterior separability and uniform posterior separability, and results on identification and recoverability in rich datasets. Overall, the two papers use different techniques and provide different insights: the two analyses can be seen as complementary. In the rest of the section, we discuss in detail similarities and differences.

As we do in Section D, Caplin, Dean, and Leahy (CDL, for short) take the perspective of an analyst who observes the behavior of an agent across finite menus $F \subset \mathbb{R}^\Theta$ and priors $\pi \in \Delta(\Theta)$. Denote by $\mathbb{D}$ the set of all decision problems (π, F) for which the analyst observes the behavior of the agent.

A first distinctive feature of CDL is that the analyst observes the behavior of the agent across *all* decision problems. To illustrate, denote by $\mathcal{F}^*$ the collection of all finite subsets of $\mathbb{R}^\Theta$. CDL assume that $\mathbb{D} = \Delta(\Theta) \times \mathcal{F}^*$. By contrast, our representation theorem (Theorems 2 and 4) apply to arbitrary finite datasets.[27] We study identification and estimation for arbitrary (finite or infinite) datasets. We assume rich datasets only in Section E to achieve unique identification.

A second distinctive feature of CDL is that the analyst observes when the agent is indifferent among alternatives. Specifically, their choice dataset is a rule that associates *a set* $S_{(\pi,F)} \subseteq \mathcal{S}_F$ of stochastic choices to each $(\pi, F) \in \Delta(\Theta) \times \mathcal{F}^*$. A cost function $C : \mathcal{M} \to [0, \infty]$ rationalizes the dataset if for every decision problem (π, F),

$$S_{(\pi,F)} = \left\{\sigma_{(\mu,d_F)} : (\mu, d_F) \text{ is a solution of (1)}\right\} .$$

If (1) has multiple solutions, then $S_{(\pi,F)}$ is not a singleton and the analyst observes when the agent is indifferent among alternatives. By contrast, our dataset assigns to each decision problem $(\pi, F) \in \mathbb{D}$ a *single* stochastic choice $\sigma_{(\pi,F)} \in \mathcal{S}_F$.

Indifferences often emerge when the cost of information is posterior separable. When

[27] They do not apply to infinite datasets, as we discuss in the context of Theorem 3.

the cost of information is posterior separable, the objective function of (1) is *affine* in μ. The affinity of the objective function leads to multiple optimal solutions in a number of settings. Matejka and McKay (2015) provide examples for the case in which H is entropy.

A third distinctive feature of CDL is that the analyst observes the agent making a wide range of choices. Let Δ_π^r be the collection of all revealed posteriors from decision problems with prior π:

$$\Delta_\pi^r = \left\{p_{(\sigma_{(\pi,F)},f)} : F \in \mathcal{F}^*, \sigma_{(\pi,F)} \in S_{(\pi,F)}, \text{ and } f \in F\right\}.$$

Let $\mathcal{M}_\pi^r$ be the collection of all distributions of revealed posteriors from decision problems with prior π:

$$\mathcal{M}_\pi^r = \left\{\mu_{\sigma_{(\pi,F)}} : F \in \mathcal{F}^* \text{ and } \sigma_{(\pi,F)} \in S_{(\pi,F)}\right\}.$$

CDL assume that, for every prior π, the analyst's dataset satisfies the following "completeness" conditions:

(i) If $p \in \Delta(\Theta)$ is such that $\operatorname{supp}(p) = \operatorname{supp}(\pi)$, then $p \in \Delta_\pi^r$.

(ii) If $\mu \in \mathcal{M}_\pi$ is such that $\operatorname{supp}(\mu) \subseteq \Delta_\pi^r$, then $\mu \in \mathcal{M}_\pi^r$.

By (i), *every* interior posterior is revealed in some decision problem. By (ii), *every* distribution over interior posteriors is revealed in some decision problem. Condition (i) and (ii) impose restrictions on the behavior of the agent that have not counterpart in our paper.

Next is the main axiom that CDL proposes for posterior separable costs:

Axiom 6. *For every $\pi \in \Delta(\Theta)$, every $F \in \mathcal{F}^*$, every $\sigma_{(\pi,F)} \in S_{(\pi,F)}$, and every $\mu \in \mathcal{M}_\pi^r$, if $\operatorname{supp}\left(\mu_{\sigma_{(\pi,F)}}\right) \cap \operatorname{supp}(\mu) \neq \varnothing$ then there exist $G \in \mathcal{F}^*$ and $\sigma_{(\pi,G)} \in S_{(\pi,G)}$ such that*

(i) $\mu_{\sigma_{(\pi,G)}} = \mu$;

(ii) for every $f \in F \cup G$ and $p \in \operatorname{supp}\left(\mu_{\sigma_{(\pi,F)}}\right) \cap \operatorname{supp}(\mu)$,

$$d_{\sigma_{(\pi,F)}}(f|p) = d_{\sigma_{(\pi,G)}}(f|p).$$

CDL interpret the axiom as a separability condition. As they put it, if two distributions of revealed posteriors have overlapping supports, then "we must be able to find decision problems that produce both distributions using common actions at shared posteriors" (Caplin, Dean and Leahy, 2017, p. 31). Overall, Axioms 3 and 6 provide complementary perspectives on posterior separability. A distinction is that Axiom 6 features an existential quantifier that may complicate testing. Axiom 6 instead has low computation complexity: checking the axiom is equivalent to solving a linear program (see Section C).

To characterize uniform posterior separability, CDL proposes the following condition, in addition to Axiom 6:

Axiom 7. *Let $\pi \in \Delta(\Theta)$, $F \in \mathcal{F}^*$, $\sigma_{(\pi,F)} \in S_{(\pi,F)}$, and $\alpha \in \Delta(F)$ such that $\alpha(f) = 0$ implies $\sum_\theta \sigma_{(\pi,F)}(f|\theta)\pi(\theta) = 0$. Define $\pi' \in \Delta(\Theta)$ by*

$$\pi'(\theta) = \sum_{f \in F} \alpha(f) p_{\left(\sigma_{(\pi,F)},f\right)}(\theta).$$

For the decision problem (π', F), consider the stochastic choice $\sigma_{(\pi',F)} : \Theta \to \Delta(F)$ given by

(i) $\sum_\theta \sigma_{(\pi',F)}(f|\theta)\pi'(\theta) = \alpha(f)$;

(ii) $\mu_{\sigma_{(\pi',F)}}(p) = \sum \left\{ \alpha(f) : p_{\left(\sigma_{(\pi,F)},f\right)} = p \right\}$;

(iii) if $p_{\left(\sigma_{(\pi,F)},f\right)} = p$, *then* $d_{\sigma_{(\pi',F)}}(f|p) = \alpha(f)/\mu_{\sigma_{(\pi',F)}}(p)$;

(iv) if $p_{\left(\sigma_{(\pi,F)},f\right)} \neq p$, *then* $d_{\sigma_{(\pi',F)}}(f|p) = 0$.

Then we must have $\sigma_{(\pi',F)} \in S_{(\pi',F)}$.

CDL interpret the axiom as an invariance condition. As they put it, the axiom "conveys the idea that, given $\sigma_{(\pi,F)} \in S_{(\pi,F)}$, the resulting action-posterior pairs are invariant to various changes in π and F" (Caplin, Dean and Leahy, 2017, p. 34). As they explain, the axiom is connected to the result *Locally Invariant Posteriors* from Caplin and Dean (2013). We use such result also in our analysis (see Lemma 4). A subtlety is that we use *Locally Invariant Posteriors* already in the context of posterior separability, even if the prior is fixed (see the discussion following Lemma 4).

Finally, CDL also provide an identification result for the cost of information in rich dataset. Their assumptions and methods are quite different from ours. First, as described above, they assume that (i) $\mathcal{F} = \mathcal{F}^*$, (ii) the analyst observes when the agent is indifferent among alternatives, and (iii) the analyst observes the agent making a wide range of choices. Under these assumptions, for every prior π, they identify the cost of information on the restricted domain $\mathcal{M}_\pi^r$. By contrast, for every prior π, we identify the cost of information on the whole domain $\mathcal{M}_\pi$, solely under the assumption that $\mathcal{F}$ is rich with respect to $\mathcal{F}^*$. CDL do not discuss test functions.

References

Aliprantis, R. and K. C. Border, *Infinite Dimensional Analysis*, 3rd ed., Springer, 2006.

Caplin, A. and M. Dean, "Behavioral implications of rational inattention with Shannon entropy," *NBER Working Paper No. 19318*, 2013.

_ , D. Csaba, J. Leahy, and O. Nov, "Rational inattention, competitive supply, and psychometrics," *The Quarterly Journal of Economics*, 2020, *135*, 1681–1724.

_ , M. Dean, and J. Leahy, "Rationally inattentive behavior: Characterizing and generalizing Shannon entropy," *NBER Working Paper No. 23652*, 2017.

Chambers, C. P. and F. Echenique, *Revealed Preference Theory*, Cambridge University Press, 2016.

_ , C. Liu, and J. Rehbeck, "Costly information acquisition," *Journal of Economic Theory*, 2020, *186.*

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

Dean, M. and N. Neligh, "Experimental tests of rational inattention," 2019. http://www.columbia.edu/ md3405/.

Denti, T., M. Marinacci, and A. Rustichini, "Experimental cost of information," *IGIER Working Paper No. 657*, 2021.

Dewan, A. and N. Neligh, "Estimating information cost functions in models of rational inattention," *Journal of Economic Theory*, 2020, *187*, 105011.

Fudenberg, D., R. Iijima, and T. Strzalecki, "Stochastic choice and revealed perturbed utility," *Econometrica*, 2015, *83*, 2371–2409.

Halevy, Y., D. Persitz, and L. Zrill, "Parametric recoverability of preferences," *Journal of Political Economy*, 2018, *126*, 1558–1593.

Lin, Y.-H., "Stochastic choice and rational inattention," 2019. https://sites.google.com/site/yihsuanyhlin/research.

Lu, J., "Random choice and private information," *Econometrica*, 2016, *84*, 1983–2027.

Matejka, F. and A. McKay, "Rational inattention to discrete choices: A new foundation for the multinomial logit model," *American Economic Review*, 2015, *105*, 272–98.

Milgrom, P. and I. Segal, "Envelope theorems for arbitrary choice sets," *Econometrica*, 2002, *70*, 583–601.

Ravid, D., "Ultimatum bargaining with rational inattention," *American Economic Review*, 2020, *100*, 2948–2963.

Rockafellar, R. T., *Convex analysis*, Princeton University Press, 1970.

Schneider, R., *Convex bodies: The Brunn–Minkowski theory*, Cambridge University Press, 2014.

Winkler, G., "Extreme points of moment sets," *Mathematics of Operations Research*, 1988, *13*, 581–587.