

Eliciting Learning Ability

Michèle Müller-Itten*

January 30, 2025

PRELIMINARY AND INCOMPLETE

Please reach out with thoughts, comments or relevant citations.

Abstract

I express an agent's learning ability through a maximal menu of zero welfare. A simple duality links this *zero set* and posterior-separable learning costs that are common in the Rational Inattention literature. The novel representation helps to elicit the agent's prior belief and perceived learning ability in a simple, cost-effective and versatile way. Relative to existing approaches, the protocol also works when learning cannot be tightly controlled, and when learning produces lasting effects that alter the beliefs or costs in subsequent tasks.

An experimental implementation is planned, and the manuscript will be updated once the findings are available.

Keywords: Rational Inattention, Learning Costs, Experimental Design

JEL: C91, D83.

*University of St.Gallen: michele.mueller-itten@unisg.ch

The author would like to thank Stefan Bucher, Samuel Häfner, Paolo Piacquadio, Leeat Yariv and participants at various conferences and seminars for constructive feedback and comments.

1 Introduction

As models of flexible, costly learning are taking on a more prominent role in economic models, the specifics of the agent’s information acquisition costs start to matter. Already, existing research shows that while the flexible learning paradigm of Rational Inattention models holds great promise in matching important patterns of behavior *qualitatively*, there is still room for improvement when it comes to *quantitative* predictions, possibly because of misspecified learning costs. In this paper, I introduce a novel approach to eliciting such costs experimentally. The method requires no control over participant’s prior knowledge, thought processes, or external resource use, making it suitable for online implementation and applicable to real-world information acquisition scenarios. An experimental validation is planned (awaiting funding approval), and will be added to this draft.

On a technical level, the key innovation lies in representing the learning technology not through a cost function that maps signals to utility losses, but through a maximal menu that yields zero welfare. This *zero set* is a collection of choice alternatives that determine how the agent’s payoff depends on the unknown state of the world, as illustrated in Figure 1(a) for a situation with two states. The agent’s prior belief regarding the state influences her initial assessment of the different alternatives, and her learning ability unlocks the possibility of tailoring her choice to the realized state, at least partially so. As such, even an optional alternative that has negative ex-ante expected utility can be welfare enhancing, so long as the agent uses learning to selectively opt either for or against that alternative. On the other hand, no payoff vector with ex-ante positive expected utility can be part of the zero set, for its unconditional selection would ensure positive welfare.

A duality links zero sets and posterior-separable learning costs. Any valid zero set is closed, convex, contains all nonpositive payoff vectors, and admits a balanced supporting hyperplane at zero.¹ A cost function is posterior-separable when it can be described as the expected value of some divergence between the posterior and the prior. Each valid zero set uniquely identifies a posterior-separable cost. Each posterior-separable cost generates a valid zero set, which is unique so long as the divergence is differentiable around the prior. For simplicity, I summarize the findings of the paper under this differentiability assumption.

¹Intuitively, a balanced supporting hyperplane means that any contact line between the set and the hyperplane is either bounded in both directions or unbounded in both.

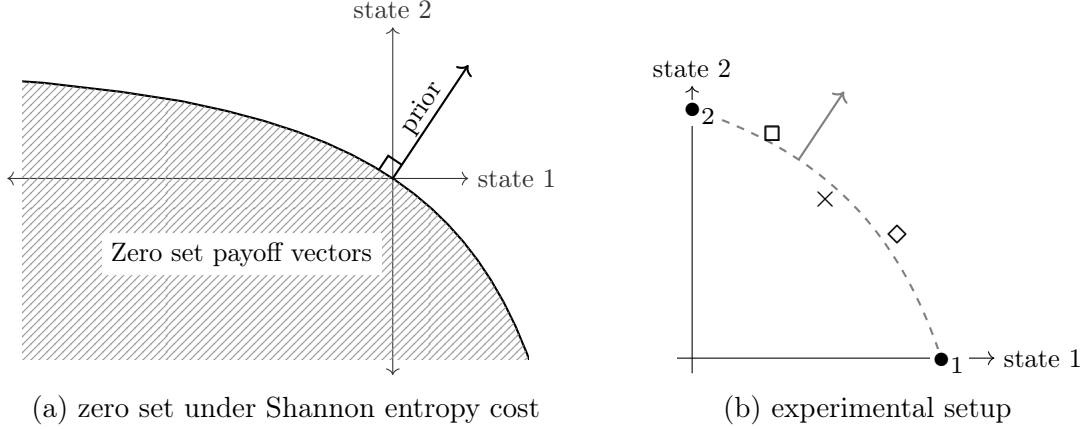


Figure 1: Visual depiction of the paper’s key concepts

The zero set not only identifies the learning cost, it also pins down the prior belief through its normal vector at the origin. Experimental identification of the zero set therefore reveals both the subject’s existing knowledge as well as her learning technology.

Two properties make the zero set particularly attractive for cost elicitation: First, despite globally identifying the agent’s knowledge and learning, it can be constructed elementwise. That is, a payoff vector is part of the zero set if and only if its presence does not improve welfare in conjunction with the zero option.² Second, when faced with the entire zero set, the agent is exactly indifferent across *all* learning strategies, for it exactly equalizes the instrumental gain and cost of any signal.

In the experiment, the agent is first asked to strike one of the alternatives in a menu of cardinality $I + 1$. After several such elimination tasks, a random number draw determines which of the trimmed menus is offered again. This time, the agent selects one alternative that determines her payout. Because the initial menus contain more alternatives than states, the agent has no incentives to acquire any information during these elimination tasks. And by recording which options are dismissed, the analyst can recover the shape of the zero set.

To illustrate the process (without proof), Figure 1(b) depicts a binary state space along with some payoff vectors. The two alternatives marked $\bullet_i$ represent a simple bet on state i : They reward the agent with a prize if and only if the world is in that state. The alternatives $\{\square, \times, \diamond\}$ represent options where the agent earns part of the

²This property echoes the menu expansion result of Müller-Itten et al. [2024], albeit for a larger class of costs.

prize in either state. Some of these options may initially be more attractive than the extreme bets, but that can change if the agent acquires conclusive information regarding one of the two states.

Concretely, the analyst might be interested in the agent’s data lookup skills, and tie the payoffs to the veracity of a claim such as: *“In the year 2007, the GDP of India is larger than that of Spain.”* State 1 payoffs apply if the claim is true, and state 2 payoffs if the claim is false. As it turns out, this claim is indeed false,³ but the agent might not know that without further inquiry.

The analyst then first asks the agent to eliminate one alternative from each of the menus

$$A_{\square} = \{\bullet_1, \bullet_2, \square\}, \quad A_{\times} = \{\bullet_1, \bullet_2, \times\}, \quad A_{\diamond} = \{\bullet_1, \bullet_2, \diamond\}$$

and later revisits one of the menus to have the agent select one of the two remaining options. Suppose that the agent eliminates alternative $\bullet_2$ from menu $A_{\square}$, alternative $\times$ from menu $A_{\times}$ and alternative $\bullet_1$ from menu $A_{\diamond}$. The analyst then traces the line that separates the payoff vectors that are themselves eliminated (here $\times$) from those that trigger the elimination of one of the extreme bets (here $\square$ and $\diamond$), as indicated by the dashed line. Next, the analyst checks the payoff vectors that are just above the gray line, and investigates where the agent changes *which* extreme bet is eliminated, as indicated by the base point of the arrow.

Once he is done, the dashed line traces what Müller-Itten et al. [2024] call the *learning-proof menu* derived from menu $\{\bullet_1, \bullet_2\}$ under the agent’s learning technology, and the base point of the arrow is the ignorance equivalent under the agent’s prior. By shifting the drawing so that the ignorance equivalent comes to lay on the origin, the analyst recovers (part of) the zero set. The section of the zero set that is uncovered grows with the size of the maximal prize that the agent can win.

While illustrated with a binary state space, the technique generalizes to any finite number of dimensions. Compared with other learning cost elicitation methods, the technique is simple and cheap to implement. After all, the absence of learning incentives means that the analyst does not need to control the agent’s access to information sources, and also does not have to generate multiple instances of the same learning scenario. And since many elimination tasks can be incentivized with a single selection and payout, larger cash prizes are feasible.

³World Bank. “World Development Indicators: GDP (current US\$).” *World Bank DataBank*, <https://databank.worldbank.org/reports.aspx?source=2&series=NY.GDP.MKTP.CD&country>. Accessed 6 Dec. 2024.

Related Literature. The quest for ‘realistic’ learning costs in the literature has spurred a formal analysis of important parametric families and experimental elicitation techniques.

Axiomatizations are available for costs that incorporate information processing constraints [Shannon, 1948, Sims, 2003], weaker forms of separability [Caplin et al., 2022, Denti, 2022], degrees of perceptual distance between states [Hébert and Woodford, 2021], linearity in the number of simultaneous experiments [Pomatto et al., 2018] or sequential information sampling [Bloedel and Zhong, 2020, Denti et al., 2022].

Experimental measurements of learning costs rely on equating the costs and benefits of information.⁴ The papers differ in how they approach the key obstacle: Learning is inherently unobservable, and the analyst never quite knows how much learning goes into each of the agent’s choices.

The analyst in de Oliveira et al. [2017] elicits the agent’s willingness to pay for menus *before* learning occurs, by offering the agent the choice between a safe payment or access to a menu of state-dependent options. The choice creates incentives for early learning, yet such learning would undermine the validity of the estimation. The authors thus restrict themselves to settings where the analyst can disable the learning technology at will, such as lookup tasks. Unfortunately, this does rule out many settings of interest, most notably those where costs arise from introspection or information processing, as those are not typically activities that the analyst can postpone to fit his schedule. It also makes it difficult to administer the method in an online setting, where the agent can easily use a secondary device to access pertinent information.⁵

The analyst in Caplin et al. [2020], Dewan and Neligh [2020] does the opposite, observing the agent’s stochastic choice from a particular menu *after* learning occurs. However, measuring state-dependent choice probabilities is only possible if the agent encounters the very same task over and over again. In the experiment, the agent assesses the number of dots or polygons on a screen, across many independently

⁴Tsakas [2020] studies a variant of this setup where the cost is known and the prior is the object of interest. In that case, agent learning is a nuisance that the analyst has to discourage in order to elicit the unbiased prior belief. Since I assume that the analyst knows neither the prior nor the cost, the new method may also be of interest in that setting.

⁵Avoyan and Romagnoli [2023] follow a similar approach, by asking the agent to select one of several symmetric menus that vary in their exposure to the underlying state. To eliminate early learning incentives, they impose symmetry assumptions on both the prior and the available learning strategies. In many real-world learning settings however, prior intuitions matter. And by restricting the class of learning strategies to a symmetric family with a single degree of freedom, their choice model loses the flexibility that is a hallmark of Rational Inattention.

generated random instances.⁶ In many situations of interest, learning is however at least partly persistent. For one, the agent may remember relevant facts learned in earlier rounds, and approach any the next task with modified beliefs. For another, she may learn to access a particular database, scan contracts more effectively, or otherwise develop expertise that lowers her learning cost for similar tasks that follow. Either way, averaging the prediction accuracy across rounds is no longer meaningful.

The present paper takes a different route by carefully avoiding incentives for learning until the very last stage of the experiment. Like in Avoyan and Romagnoli [2023], de Oliveira et al. [2017], my analyst repeatedly offers the agent a choice over menus, with the promise that he will later get to select one of the options in the chosen menu. But rather than comparing a rich menu with a safe option, our analyst offers choices between several menus that are each rich enough to undercut any benefits from early learning. As a consequence, the protocol can be applied to elicit perceived learning costs in settings that are inaccessible to existing methods. The learning costs could stem from a cognitive effort required to parse salient details from insurance contracts or to process questions on standardized tests. They could measure tech-savviness that affects the ease of obtaining financial information. Or they could measure access to information sources that are unobserved by the analyst.

Paper structure. Section 2 establishes the duality between posterior-separable cost functions and the zero set, and how the zero set can be identified from (idealized) experimental data. Section 3 is pending and will report on the findings from a planned online implementation of the protocol. All proofs are in the appendix; hyperlinks connect each theorem to the corresponding proof.

2 Theoretical Basis

2.1 Decision Model

An agent chooses one action from a **menu** A while facing an a priori unknown **state** of nature $i \in \mathcal{I} = \{1, \dots, I\}$. Actions differ only in their state-dependent payouts, allowing us to identify an action directly through its state-dependent payouts $\mathbf{a} \in \mathbb{R}^{\mathcal{I}}$.

The agent is risk-neutral and privately endowed with a **prior belief** $\pi \in \Delta\mathcal{I}$.

⁶Caplin et al. [2020] argues that this setup is similar to ‘simple production tasks’, such as identifying defective items on a manufacturing line.

For maximum generality of the theoretical results, I do not initially assume that the prior has full support.

Prior to choosing an action, the agent can learn. Learning refines her belief according to Bayes' rule. As is common in the Rational Inattention (RI) literature, I model learning as a distribution over posterior beliefs that averages to the prior. Let $\Delta\Delta\mathcal{I} \mid \pi := \{F \in \Delta\Delta\mathcal{I} \mid \mathbb{E}[\boldsymbol{\rho} \mid \boldsymbol{\rho} \sim F] = \pi\}$ denote the set of all such Bayes-plausible belief distributions.

Learning is costly. The **learning cost** function $c^\pi : \Delta\Delta\mathcal{I} \mid \pi \rightarrow \mathbb{R} \cup \{\infty\}$ captures how costly it is to follow a particular learning strategy; $c^\pi(F)$ utils are subtracted from the agent's expected consumption utility if she learns according to F . The cost is posterior separable, so it can be expressed as the expected increase of some (prior-dependent) **potential** $\hat{\phi}^\pi : \Delta\mathcal{I} \rightarrow \mathbb{R} \cup \{\infty\}$,

$$c^\pi(F) = \mathbb{E} \left[\hat{\phi}^\pi(\boldsymbol{\rho}) \mid \boldsymbol{\rho} \sim F \right] - \hat{\phi}^\pi(\pi). \quad (1)$$

The potential is defined on the extended reals, allowing $\hat{\phi}^\pi(\boldsymbol{\rho}) = \infty$ to delineate some posteriors as impossible to reach. It also satisfies a few standard properties:

- (A1) The potential is *finite at the prior*, which is necessary for costs (1) to be well defined.
- (A2) The potential is *convex*, capturing the fact that learning costs are monotonic in the Blackwell order. Mathematically, convexity of the potential is equivalent to its epigraph

$$\text{epi}(\hat{\phi}^\pi) := \left\{ (\boldsymbol{\sigma}, r) \in \Delta\mathcal{I} \times \mathbb{R} \mid \hat{\phi}^\pi(\boldsymbol{\sigma}) \leq r \right\}, \quad (2)$$

the area above its graph in $\Delta\mathcal{I} \times \mathbb{R}$, being convex.

- (A3) The potential is *lower semicontinuous*,⁷ reflecting the fact that the potential does not suddenly jump up at the limit of a finite-valued sequence. Mathematically, lower semicontinuity of $\hat{\phi}^\pi$ is equivalent to its epigraph being closed.⁸ Lower semicontinuity is not a very strong assumption: Convexity already implies continuity on the interior of $\text{dom}(\hat{\phi}^\pi) = \{\boldsymbol{\rho}^j \in \Delta\mathcal{I} \mid \hat{\phi}^\pi(\boldsymbol{\rho}^j) < \infty\}$, the effective domain of $\hat{\phi}^\pi$.⁹

⁷The function $\hat{\phi}^\pi$ is lower semicontinuous at $\boldsymbol{\rho}^0$ if and only if $\liminf_{\boldsymbol{\rho}^n \rightarrow \boldsymbol{\rho}^0} \hat{\phi}^\pi(\boldsymbol{\rho}^n) \geq \hat{\phi}^\pi(\boldsymbol{\rho}^0)$.

⁸See, e.g. Proposition 1.1.2 in Bertsekas [2009].

⁹See, e.g. Proposition 1.3.11 in Bertsekas [2009].

The assumptions are weaker than those of Caplin et al. [2022] in two important regards: By requiring only weak convexity, I allow the agent to encounter free gains in additional precision, and by allowing the potential to be infinite in the interior of the simplex, the model can accommodate firm information processing constraints.

Normalization. Although it is convenient to express costs in terms of the expected gain in potential, there are two ways in which Equation (1) introduces spurious degrees of freedom into the characterization of the learning technology.

First, suppose that any belief update in the direction of $\boldsymbol{\rho} \in \Delta\mathcal{I}$ engenders infinite cost, $\hat{\phi}^\pi(\boldsymbol{\pi} + \varepsilon(\boldsymbol{\rho} - \boldsymbol{\pi})) = \infty$ for all $\varepsilon > 0$. Since the average posterior equals the prior, this rules out all learning strategies with a belief update in this or the opposite direction. It is thus without loss of generality to also assume that $\hat{\phi}^\pi(\boldsymbol{\pi} - \varepsilon(\boldsymbol{\rho} - \boldsymbol{\pi})) = \infty$ for all $\varepsilon > 0$.

Second, adding any affine function $\boldsymbol{\rho} \mapsto \mathbf{v} \cdot \boldsymbol{\rho} + m$, with $\mathbf{v} \in \mathbb{R}^{\mathcal{I}}$ and $m \in \mathbb{R}$, to $\hat{\phi}^\pi$ does not alter the resulting cost c^π . All terms cancel out since $\mathbb{E}[\mathbf{v} \cdot \boldsymbol{\rho} + m \mid \boldsymbol{\rho} \sim F] = \mathbf{v} \cdot \boldsymbol{\pi} + m$ for any distribution in $\Delta\Delta\mathcal{I} \mid \boldsymbol{\pi}$.

To state the main results more succinctly, it is useful to normalize the potential so that it satisfies two additional properties:

(A4) The prior $\boldsymbol{\pi}$ is in the relative interior of the potential’s effective domain.¹⁰

(A5) The set $\{(\boldsymbol{\rho}, 0) \mid \boldsymbol{\rho} \in \Delta\mathcal{I}\}$ forms a supporting hyperplane of the potential at the prior $\boldsymbol{\pi}$.

In particular, Assumption A5 normalizes the potential of the prior to zero.

For later results, it turns out to be useful to relate these assumptions to the normal cone of $\text{epi}(\hat{\phi}^\pi)$ at the prior, denoted

$$N_\pi(\hat{\phi}^\pi) = \left\{ \begin{bmatrix} \mathbf{n} \\ n_0 \end{bmatrix} \in \mathbb{R}^{\mathcal{I}} \times \mathbb{R} \mid \begin{bmatrix} \mathbf{n} \\ n_0 \end{bmatrix} \cdot \left(\begin{bmatrix} \boldsymbol{\rho} \\ s \end{bmatrix} - \begin{bmatrix} \boldsymbol{\pi} \\ \hat{\phi}^\pi(\boldsymbol{\pi}) \end{bmatrix} \right) \leq 0 \text{ for all } \begin{bmatrix} \boldsymbol{\rho} \\ s \end{bmatrix} \in \text{epi}(\hat{\phi}^\pi) \right\}.$$

By known results in convex analysis,¹¹ Assumption A4 implies in particular that

¹⁰By Assumption A2, the effective domain is convex, which allows us to define the relative interior using the Prolongation Lemma [Proposition 1.3.3 in Bertsekas, 2009]. Indeed, the prior is in the relative interior of $\text{dom}(\hat{\phi}^\pi)$ whenever, for all $\boldsymbol{\rho} \in \text{dom}(\hat{\phi}^\pi)$, the line segment from $\boldsymbol{\rho}$ to $\boldsymbol{\pi}$ can be prolonged beyond $\boldsymbol{\pi}$ without leaving the $\text{dom}(\hat{\phi}^\pi)$, i.e. $\exists \varepsilon > 0$ such that $\boldsymbol{\pi} + \varepsilon(\boldsymbol{\pi} - \boldsymbol{\rho}) \in \text{dom}(\hat{\phi}^\pi)$.

¹¹Spelling out the inequality for any $(\mathbf{v}, -1) \in N_\pi(\hat{\phi}^\pi)$ yields the subgradient equation at $\boldsymbol{\pi}$. The existence of a subgradient for a point of $\text{rel.int}(\text{dom}(\hat{\phi}^\pi))$ is shown, e.g., in Proposition 5.4.1 of Bertsekas [2009].

$N_{\pi}(\phi^{\pi})$ contains some vector $(\mathbf{v}, -1)$, while Assumption A5 implies that the vector $(\mathbf{0}, -1)$ is contained in $N_{\pi}(\phi^{\pi})$.

Lemma 1 provides a closed form expression for the normalization and ensures that it is without loss of generality. In what follows, I therefore restrict attention to normalized potentials, and write ϕ^{π} rather than $\hat{\phi}^{\pi}$.

Lemma 1 (Normalization). *Consider a potential $\hat{\phi}^{\pi} : \Delta\mathcal{I} \rightarrow \mathbb{R} \cup \{\infty\}$ with associated cost $c^{\pi}(F) = \mathbb{E} [\hat{\phi}^{\pi}(\boldsymbol{\rho}) \mid \boldsymbol{\rho} \sim F] - \hat{\phi}^{\pi}(\boldsymbol{\pi})$.*

(a) *If $\hat{\phi}^{\pi}$ satisfies Assumptions A1 to A3, then $\phi^{\pi} : \Delta\mathcal{I} \rightarrow \mathbb{R} \cup \{\infty\}$ with*

$$\phi^{\pi}(\boldsymbol{\rho}) := \begin{cases} \hat{\phi}^{\pi}(\boldsymbol{\rho}) & \text{if } \boldsymbol{\pi} + t(\boldsymbol{\pi} - \boldsymbol{\rho}) \in \text{dom}(\hat{\phi}^{\pi}) \text{ for some } t > 0 \\ \infty & \text{otherwise} \end{cases}$$

satisfies Assumptions A1 to A4.

(b) *If $\hat{\phi}^{\pi}$ satisfies Assumptions A1 to A4, then $N_{\pi}(\hat{\phi}^{\pi})$ contains at a vector $(\mathbf{v}, -1) \in \mathbb{R}^{\mathcal{I}} \times \mathbb{R}$. Moreover, $\phi^{\pi} : \Delta\mathcal{I} \rightarrow \mathbb{R} \cup \{\infty\}$ with*

$$\phi^{\pi}(\boldsymbol{\rho}) := \hat{\phi}^{\pi}(\boldsymbol{\rho}) + \mathbf{v} \cdot (\boldsymbol{\pi} - \boldsymbol{\rho}) - \hat{\phi}^{\pi}(\boldsymbol{\pi})$$

satisfies Assumptions A1 to A5.

In either case, ϕ^{π} maintains the cost function,

$$\mathbb{E} [\phi^{\pi}(\boldsymbol{\rho}) \mid \boldsymbol{\rho} \sim F] - \phi^{\pi}(\boldsymbol{\pi}) = c^{\pi}(F) \quad \forall F \in \Delta\Delta\mathcal{I} \mid_{\pi}.$$

Proof. See Appendix A.1. □

Optimal behavior. The agent's welfare from a particular menu A is given by

$$U(A \mid \boldsymbol{\pi}, c^{\pi}) := \sup_{F \in \Delta\Delta\mathcal{I} \mid_{\pi}} \mathbb{E} \left[\left(\sup_{\mathbf{a} \in A} \boldsymbol{\rho} \cdot \mathbf{a} \right) - \phi^{\pi}(\boldsymbol{\rho}) \mid \boldsymbol{\rho} \sim F \right]. \quad (3)$$

Lower semi-continuity of the cost potential implies that the maximum is attained for any nonempty and closed menu A that is bounded above, so the agent indeed has an optimal selection strategy (or several).

By linearity of the expectation, it also follows that adding the same payoff vector $\mathbf{v}$ to every alternative in the menu increases consumption utility by $\mathbb{E} [\boldsymbol{\rho} \cdot \mathbf{v} \mid \boldsymbol{\rho} \sim F] \equiv$

$\boldsymbol{\pi} \cdot \boldsymbol{v}$ irrespective of the strategy $F \in \Delta\Delta\mathcal{I} \mid_{\boldsymbol{\pi}}$, and hence welfare satisfies

$$U(A + \{\boldsymbol{v}\} \mid \boldsymbol{\pi}, c^{\boldsymbol{\pi}}) \equiv U(A \mid \boldsymbol{\pi}, c^{\boldsymbol{\pi}}) + \boldsymbol{\pi} \cdot \boldsymbol{v} \quad \forall \boldsymbol{v} \in \mathbb{R}^{\mathcal{I}}. \quad (4)$$

Uniqueness of the normalized potential. Intuitively speaking, the normalized potential is unique so long as there are no finite fixed costs to learning. In terms of behavior, finite fixed costs are made manifest when the agent is choosing between two payoff vectors that differ only by some multiple of a vector with zero expected utility. If the agent breaks the tie through learning only when the scaling constant is large enough, she is facing finite fixed costs in that direction that need to be outweighed by potential consumption gains.

Lemma 2 formalizes this argument, first by characterizing the relation between any two equivalent normalized potentials, and then relating uniqueness to the welfare from a binary menu with two payoff vectors of zero expected utility.

Lemma 2. *Normalized potentials $\phi^{\boldsymbol{\pi}}$ and $\psi^{\boldsymbol{\pi}}$ represent the same cost if and only if there exists $(\boldsymbol{v}, -1) \in N_{\boldsymbol{\pi}}(\phi^{\boldsymbol{\pi}})$ such that $\psi^{\boldsymbol{\pi}}(\boldsymbol{\rho}) \equiv \phi^{\boldsymbol{\pi}}(\boldsymbol{\rho}) + \boldsymbol{v} \cdot (\boldsymbol{\pi} - \boldsymbol{\rho})$.*

Moreover, the following are equivalent:

- (a) *The normalized cost potential is unique.*
- (b) *$\boldsymbol{v} \cdot (\boldsymbol{\pi} - \boldsymbol{\rho}) = 0$ for any $(\boldsymbol{v}, -1) \in N_{\boldsymbol{\pi}}(\phi^{\boldsymbol{\pi}})$ and any $\boldsymbol{\rho} \in \text{dom}(\phi^{\boldsymbol{\pi}})$.*
- (c) *For all $\boldsymbol{\rho} \in \text{dom}(\phi^{\boldsymbol{\pi}})$ and $\delta > 0$, $\exists \varepsilon \in (0, 1)$ such that $\phi^{\boldsymbol{\pi}}(\boldsymbol{\pi} + \varepsilon(\boldsymbol{\rho} - \boldsymbol{\pi})) < \delta\varepsilon$.*
- (d) *For all $\boldsymbol{a} \perp \boldsymbol{\pi}$, the sign of $U(\{\mathbf{0}, k\boldsymbol{a}\} \mid \boldsymbol{\pi}, c^{\boldsymbol{\pi}})$ is the same across all $k > 0$.*

Proof. See Appendix A.1. □

In the special case where costs can be represented by some finite potential $\hat{\phi}^{\boldsymbol{\pi}}$ (not necessarily normalized) and the prior $\boldsymbol{\pi}$ has full support, uniqueness of the normalized potential is equivalent to differentiability of $\hat{\phi}^{\boldsymbol{\pi}}$ at the prior. After all, this ensures that the subgradient is unique. The normalization procedure from Lemma 1 then yields a normalized potential $\phi^{\boldsymbol{\pi}}$ where $\boldsymbol{v} = \mathbf{0}$ is the only vector with $(\boldsymbol{v}, -1) \in N_{\boldsymbol{\pi}}(\phi^{\boldsymbol{\pi}})$.

2.2 Zero set

The point of departure from the standard setup lies in a duality between the agent's cost potential $\phi^{\boldsymbol{\pi}}$ and a maximal payoff set that offers the agent zero welfare.

Definition. I assign a payoff vector to each hyperplane that lies entirely below the normalized potential ϕ^π by evaluating its value at the extremal points of the simplex,

$$Z(\phi^\pi) := \{z \in \mathbb{R}^I \mid \phi^\pi(\rho) \geq z \cdot \rho \ \forall \rho \in \Delta\mathcal{I}\}. \quad (5)$$

I refer to this as the **zero set**, for is a maximal set of zero welfare.

Proposition 1. $U(Z(\phi^\pi) \mid \pi, c^\pi) = 0 < U(Z(\phi^\pi) \cup \{a\} \mid \pi, c^\pi)$ for all $a \notin Z(\phi^\pi)$.

Proof. See Appendix A.2. □

Since the hyperplanes below ϕ^π form a closed and convex set, the same is true of the zero set. And since each of these hyperplanes can be arbitrarily lowered, the zero set contains the convex cone $\mathbb{R}_{\leq 0}^I$. Finally, the definition requires in particular that $\phi^\pi(\pi) = 0 \geq z \cdot \pi$ for all $z \in Z$, hence the prior describes a supporting hyperplane at the origin. This is intuitive: The zero set cannot include any payoff vector with positive expected return, lest its unconditional implementation already exceeds welfare 0.

The normalization of the potential implies something slightly stronger, albeit a bit technical: I say that a vector $v \in \mathbb{R}^I$ yields a *balanced* supporting hyperplane $H = \{z \in \mathbb{R}^I \mid z \cdot v = 0\}$ of Z at the origin whenever $z \cdot v \leq 0$ for all $z \in Z$ and the intersection $H \cap Z$ can be written as the Minkowski sum of a nonempty compact set and a linear subspace. Balanced support means that all maximal straight line segments of Z that are “supported by” (i.e. part of) H are either unbounded in both directions or bounded in both. Figure 2 shows an example where this condition is violated, since $Z \cap H$ (the thick blue line) is unbounded only in one direction. Proposition 2 below shows that the prior forms such a balanced supporting hyperplane at the origin.

For an analyst, the zero set is helpful because it uniquely identifies the prior and the cost function. The prior belief is orthogonal to the zero set at the origin, and the cost potential $\phi^\pi(\rho)$ is obtained simply by maximizing expected utility over the zero set in direction ρ .

Proposition 2. *The zero set has the following properties:*

(a) *The prior is orthogonal to the zero set at the origin, $\pi \cdot z \leq 0 \ \forall z \in Z(\phi^\pi)$.*

More specifically, the prior yields a balanced supporting hyperplane of $Z(\phi^\pi)$.

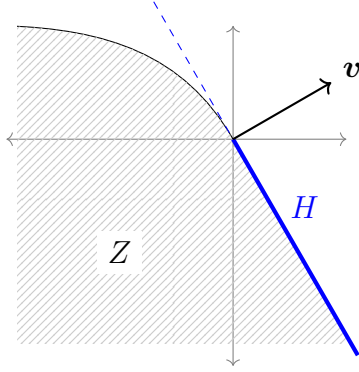


Figure 2: unbalanced H

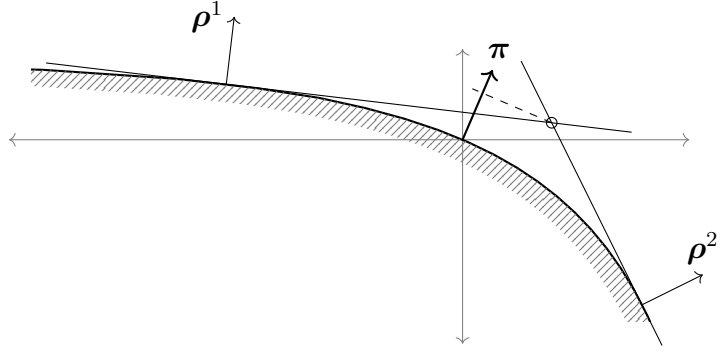


Figure 3: Zero set and the cost of a binary signal

(b) Maximizing expected utility in any direction $\rho \in \Delta\mathcal{I}$ yields the cost potential, $\sup_{z \in Z(\phi^\pi)} \rho \cdot z = \phi^\pi(\rho)$.

When ϕ^π admits a subgradient at ρ , the supremum is attained.

Proof. See Appendix A.2. □

As a direct corollary to Claim (b), the agent is not only willing to unconditionally implement $\mathbf{0}$ when given access to the zero set – in fact, she is indifferent across *all* feasible learning strategies. They all yield welfare zero.

Corollary 1. $\mathbb{E} [\sup_{z \in Z(\phi^\pi)} \rho \cdot z \mid \rho \sim F] = c^\pi(F)$ for any strategy $F \in \Delta\Delta\mathcal{I} \mid_\pi$.

Put differently, the zero set simultaneously equates costs and benefits from all learning strategies. Consequently, a signal’s cost can be ‘read’ from the zero set by taking tangents at the resulting posteriors, and evaluating the expected utility of the intersection point under the prior belief, as illustrated in Figure 3 for the case of a binary signal. The geometric construction generalizes to any signal where the signal realization space has at most the same cardinality as the state space: All tangent planes necessarily intersect, and all points in the intersection of the tangent planes have the same expected utility under the prior.¹² For richer signals, mathematical averaging is required.

Set Properties. Since the pointwise maximum is always convex, Proposition 2 also implies that the zero set can itself be taken as the primitive to describe an

¹²Formally, expected utility is equal to $\pi \cdot x = \mathbb{E}[\rho \mid \rho \sim F] \cdot x = \mathbb{E}[\rho \cdot x \mid \rho \sim F] = \mathbb{E}[\phi^\pi(\rho) \mid \rho \sim F] = c^\pi(F)$ for any x in the intersection.

information technology. Put differently, whenever a set Z is closed, convex, contains all nonpositive vectors and admits a balanced supporting hyperplane at the origin, it corresponds to the zero set of some RI agent. In that sense, these properties of the zero set are *sufficient* to fully characterize the agent’s belief and learning ability.

Proposition 3. *Consider any set $Z \in \mathbb{R}^{\mathcal{I}}$ that is closed, convex, contains the convex cone $\mathbb{R}_{\leq 0}^{\mathcal{I}}$, and admits a nonzero vector $\mathbf{v} \in \mathbb{R}^{\mathcal{I}}$ that yields a balanced supporting hyperplane at the origin. Then $Z = Z(\phi^Z)$ for an RI agent with prior $\boldsymbol{\pi}^Z := \frac{1}{\mathbf{1} \cdot \mathbf{v}} \mathbf{v} \in \Delta \mathcal{I}$ and normalized cost potential $\phi^Z(\boldsymbol{\rho}) := \sup_{\mathbf{z} \in Z} \boldsymbol{\rho} \cdot \mathbf{z}$.*

Proof. See Appendix A.2. □

Special Cases. Costs are *uniformly posterior separable* when the same cost potential $\hat{\phi}$ generates the cost independent of the prior belief $\boldsymbol{\pi} \in \text{rel.int}(\text{dom}(\hat{\phi}))$. Normalization still introduces some prior-dependency, but in a very structured way that allows us to easily track the implications for the zero set. Indeed, costs are uniformly separable if and only if a change in the agent’s prior additively translates the zero set in space.

Proposition 4. *Let $\hat{\phi} : \Delta \mathcal{I} \rightarrow \mathbb{R} \cup \{\infty\}$ satisfy Assumptions A2 and A3. If the learning costs under priors $\boldsymbol{\pi}^1, \boldsymbol{\pi}^2 \in \text{rel.int}(\text{dom}(\hat{\phi}))$ are both generated by $\hat{\phi}$, then there exists $\mathbf{z}^0 \in \arg \max \{\boldsymbol{\pi}^2 \cdot \mathbf{z} \mid \mathbf{z} \in Z(\phi^1)\}$ such that*

$$Z(\phi^2) = Z(\phi^1) - \{\mathbf{z}^0\},$$

where each ϕ^i represents a normalized potential under prior $\boldsymbol{\pi}^i$.

Proof. See Appendix A.2. □

To provide visual intuition for how to ‘read’ zero sets, Figure 4 plots $Z(c^\pi)$ for various cost functions in the binary state case under the common prior $\boldsymbol{\pi}$. Free (c^0) and infinitely costly learning (c^∞) serve as benchmarks. The cost c^r represents an information technology that grants the decision maker free access to a symmetric signal of precision $r = 0.8$ (and any garblings thereof). This can also be interpreted as a hard information processing constraint, as in Sims [2003]. In the zero set, this results in a cone whose sides are perpendicular to the agent’s posteriors after observing the free signal. Costs c^d and c^s both represent flexible learning technologies where the cost of a signal is increasing in its precision. The agent with cost c^s can informally be

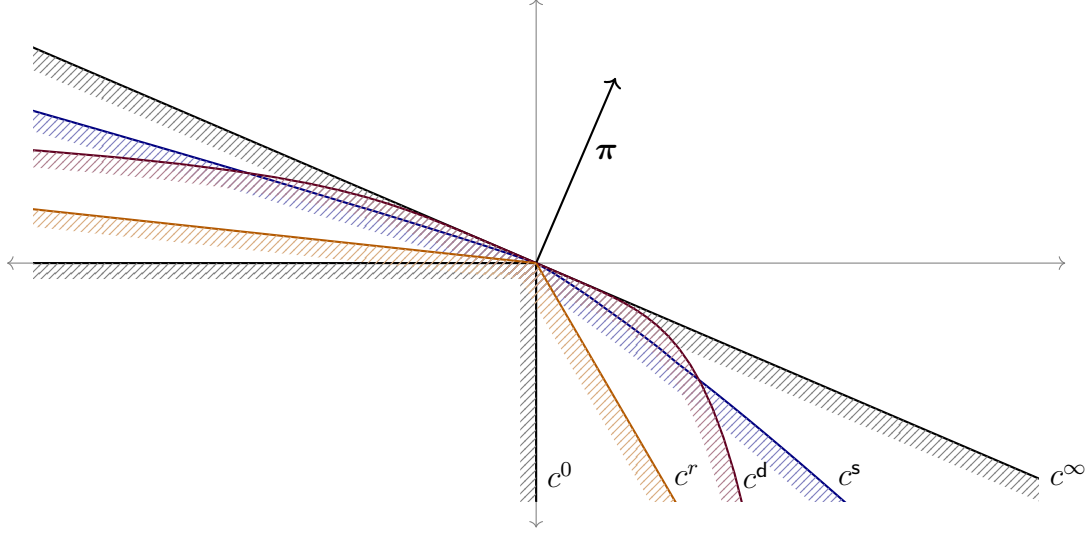


Figure 4: Zero sets for various cost functions

described as a ‘shallow learner’ and the agent with cost c^d is a ‘deep learner’, for the following reason: When these agents have to come up with a rough first guess that moves the prior only little, the shallow learner has a cost advantage that is indicated by the comparatively higher curvature of his zero set at the origin. Conversely, when the agents have to carry out an accurate assessment that results in near degenerate posteriors, the deep learner has a cost advantage that results from the higher overall curvature of his zero set when we zoom away from the origin.

Costs can also be asymmetric, which would correspond to a zero set whose boundary is close to the c^∞ line in the upper left quadrant, and closer to the c^0 line in the lower right quadrant. This represents a situation where it is relatively easy for the agent to find evidence that conclusively points to state 1, but difficult to confirm the opposite state.

Of course, zero sets are also well defined in multiple dimensions. It is entirely possible for an agent to distinguish between some states with ease, all while having difficulty to tell apart some other states. Closed form parametric expressions for the zero sets of commonly used cost functions from the literature are provided in Appendix B.1.

Uniqueness of the zero set. The cost-preserving affine transformations from Lemma 2 correspond to a vector translation of the zero set by some $\mathbf{z} \in Z(\phi^\pi)$

with zero expected utility.¹³ When the normalized potential fails to be unique, these translations meaningfully shift the zero set.¹⁴ When the normalized cost potential is unique, the part of the zero set with zero expected utility, $Z^0 = \{z \in Z(c^\pi) \mid z \cdot \pi = 0\}$, forms a linear subspace. As such, it is already closed under addition, and the translations are inconsequential.

In what follows, I assume that the normalized cost potential, and hence the zero set, is unique. I thus refer to the zero set interchangeably as $Z(c^\pi)$ or $Z(\phi^\pi)$.

(A6) The normalized cost potential is unique.

Contrary to Assumptions A4 and A5, this is a substantive restriction on the set of admissible cost functions, and its behavioral implication are made explicit in Lemma 2(d).

Binary characterization. Relative to the cost potential, the advantage of the zero set is that it can be constructed element by element. Indeed, a payoff vector is part of the zero set if and only if it would be dismissed outright against the zero outside option.

Proposition 5. $Z(c^\pi) = \{z \in \mathbb{R}^I \mid U(\{z, 0\} \mid \pi, c^\pi) \leq 0\}$.

Proof. See Appendix A.2. □

Without Assumption A6, this equivalence breaks down. In this case, the union of payoff vectors that are individually dismissed outright is equal to $Z(c^\pi) + Z^0$, which is strictly larger than $Z(c^\pi)$ and yields positive welfare (by Proposition 1).

Experimentally, the ‘pointwise’ identification of the learning technology is easier to handle than the ‘global’ concept of the cost potential. Still, there remain a few practical issues: In the presence of noise, identifying the agent’s first point of indifference between two menus is tricky. The bias of even a small imprecision can be substantial. And since subject payments are bounded below, negative payouts are not feasible. Section 2.3 deals with these issues.

¹³Formally, for any $(v, -1) \in N_\pi(\phi^\pi)$, we have $\psi^\pi(\rho) := \phi^\pi(\rho) + v \cdot (\pi - \rho) \leq z \cdot \rho$ if and only if $\phi^\pi(\rho) \leq (z + v - (v \cdot \pi)\mathbf{1}) \cdot \rho$. Thus, $Z(\psi^\pi) = Z(\phi^\pi) + \{z^0\}$ for $z^0 := v - (v \cdot \pi)\mathbf{1}$ with $z^0 \cdot \pi = 0$. Further, $(v, -1) \in N_\pi(\phi^\pi)$ implies that $v \cdot (\rho - \pi) \leq \phi^\pi(\rho)$ for all $\rho \in \Delta\mathcal{I}$ and hence $z^0 \in Z(\phi^\pi)$.

¹⁴Connecting Footnote 13 and Lemma 2(b), note that uniqueness of the normalized potential is equivalent to $z^0 \cdot \rho \equiv 0$ over the entire effective domain, and hence $z \pm z^0$ is already contained in $Z(\phi^\pi)$.

Relation to existing approaches. The ‘posterior-based’ approach [Caplin et al., 2022] is most commonly used to approach rational inattention problems in the literature. Visually, this approach maps posteriors $\boldsymbol{\rho}$ to an expected ‘net utility’ for each available payoff vector $\mathbf{a}$ by computing $\boldsymbol{\rho} \cdot \mathbf{a} - \phi^\pi(\boldsymbol{\rho})$ and then considers the concave upper envelope across all $\mathbf{a} \in A$. The optimal posteriors are then determined by the tangency points with the supporting hyperplane at the prior. Formally, the supporting hyperplane is a linear function $\boldsymbol{\rho} \mapsto \boldsymbol{\alpha} \cdot \boldsymbol{\rho}$ with

$$\left[\sup_{\mathbf{a} \in A} \boldsymbol{\rho} \cdot \mathbf{a} \right] - \phi^\pi(\boldsymbol{\rho}) \leq \boldsymbol{\alpha} \cdot \boldsymbol{\rho} \quad \forall \boldsymbol{\rho} \in \Delta\mathcal{I},$$

such that the inequality binds for all posteriors $\boldsymbol{\rho}$ in the support of the optimal strategy. In particular, $U(A \mid \boldsymbol{\pi}, c^\pi) = \boldsymbol{\alpha} \cdot \boldsymbol{\pi}$.

Restating the inequality and applying Proposition 2(b) yields

$$\boldsymbol{\rho} \cdot \mathbf{a} \leq \boldsymbol{\alpha} \cdot \boldsymbol{\rho} + \phi^\pi(\boldsymbol{\rho}) = \sup \{ \boldsymbol{\rho} \cdot (\mathbf{z} + \boldsymbol{\alpha}) \mid \mathbf{z} \in Z(c^\pi) \} \quad \forall \boldsymbol{\rho} \in \Delta\mathcal{I}, \forall \mathbf{a} \in A.$$

In other words, there exists a payoff vector $\boldsymbol{\alpha} \in \mathbb{R}^{\mathcal{I}}$ with expected utility $U(A \mid \boldsymbol{\pi}, c^\pi)$ such that the menu A is entirely contained in $Z(c^\pi) + \{\boldsymbol{\alpha}\}$. This payoff vector is exactly the ignorance equivalent of Müller-Itten et al. [2024]. After all, the agent is indifferent between the original menu, the singleton $\{\boldsymbol{\alpha}\}$, and their union $A \cup \{\boldsymbol{\alpha}\}$.

$$U(A \mid \boldsymbol{\pi}, c^\pi) = U(\{\boldsymbol{\alpha}\} \mid \boldsymbol{\pi}, c^\pi) = U(A \cup \{\boldsymbol{\alpha}\} \mid \boldsymbol{\pi}, c^\pi).$$

The upper concave envelope over the net utilities in all three problems is bounded by the same hyperplane, which yields $\boldsymbol{\pi} \cdot \boldsymbol{\alpha}$ at the prior. Thus, locating the ignorance equivalent of menu A under prior $\boldsymbol{\pi}$ and learning cost c^π requires solving the constrained optimization problem

$$\begin{cases} \min_{\boldsymbol{\alpha} \in \mathbb{R}^{\mathcal{I}}} & \boldsymbol{\pi} \cdot \boldsymbol{\alpha} \\ \text{s.t.} & A - \{\boldsymbol{\alpha}\} \subseteq Z(c^\pi). \end{cases}$$

When the zero set is taken as the primitive of the learning technology, it is thus possible to sidestep the optimal learning strategy when locating the ignorance equivalent.

2.3 Experimental Design

In the archetypal choice task, the agent selects one alternative from a menu and enjoys the associated state-dependent payoffs. In the experiment however, the agent identifies the learning technology through by repeatedly asking the agent to dismiss an alternative from various menus. For clarity, I refer to the former as a *selection task* and the latter as an *elimination task*. I say that an alternative $\mathbf{a} \in A$ is *critical* when $U(A \mid \boldsymbol{\pi}, c^\pi) > U(A \setminus \{\mathbf{a}\} \mid \boldsymbol{\pi}, c^\pi)$, so that it contributes positively to agent welfare in the selection task.

The purpose of this section is to show how the analyst can back out (part of) the zero set boundary from an appropriate sequence of elimination tasks, all while discouraging on-path learning.

Incentivizing elimination tasks. To be meaningful, the elimination choices need to carry payoff consequences. One way to incentivize the dismissal of alternative $\mathbf{d} \in A$ is to later confront the agent with the selection task from the trimmed menu $A \setminus \{\mathbf{d}\}$. Another way is to stochastically incentivize multiple elimination tasks through a single selection task. That is, after the agent eliminates option $\mathbf{d}^k$ from menu A^k for several k , a random draw determines which of the menus $A^k \setminus \{\mathbf{d}^k\}$ is offered for selection. Importantly, the agent cannot anticipate or influence this random draw, not even at a cost: The random draw is independent of the state and independent of the agent’s behavior.

The stochastic incentivization serves two important purposes: First, the selection task affects the subject’s payout and thus is costly for the experimenter. Larger stakes are feasible when the experimenter can incentivize a whole series of elimination tasks with a single selection task. Second, uncertainty about payoff-relevance, however small, encourages the agent to postpone learning until the moment where the information becomes actionable. Indeed, if the elimination task is surely followed by selection, the two task simply collapse to a selection from the entire menu A . It is then always optimal for the agent to immediately acquire the optimal selection information, determine a posterior-optimal alternative $\mathbf{a}$ and dismiss any of the non-optimal options in $A \setminus \{\mathbf{a}\}$. This is terrible for the analyst who cannot observe the counterfactual realizations: All he can conclude from the agent’s dismissal of option $\mathbf{d} \in A$ is that an unconditional choice of $\mathbf{d}$ is not the uniquely optimal selection from that menu. With stochastic incentivization, the agent first assesses whether the menu contains any non-critical alternative. If one exists, she dismisses it outright

and waits to see whether it pays to actually engage in costly learning. Now, *if* the analyst knows that the menu contains a non-critical alternative, observing dismissal of alternative $\mathbf{d} \in A$ means that $\mathbf{d}$ is not part of *any* optimal selection strategy, which is much stronger. The absence of early learning also implies that the agent’s prior is maintained across all elimination tasks. Only once the selection is due, the agent (potentially) learns and adapts her beliefs.

Designing elimination tasks. I assume that the analyst has access to a ‘base menu’ $B = \{\mathbf{b}^1, \dots, \mathbf{b}^I\} \subset \mathbb{R}_{\geq 0}^I$ of feasible payout vectors that are nonredundant in so far as each contributes positively to agent welfare in the selection task,

(A7) All I alternatives in the base menu B are critical.

Assumption A7 is really a statement about agent learning being sufficiently flexible, so that it pays to exploit the distinct payoff combinations of each alternative.

I also assume that the zero set is smooth at all points except possibly those with zero expected utility,

(A8) There is a unique normal vector $\boldsymbol{\sigma} \in \Delta\mathcal{I}$ at each $\mathbf{z} \in \partial Z(c^\pi)$ with $\boldsymbol{\pi} \cdot \mathbf{z} < 0$.

In light of Proposition 3, I chose to state this property directly in terms of the zero set, using this as the primitive of the learning technology. In terms of cost, it implies that there is no ‘atomicity’ to learning, except perhaps some initial information that is completely free.

While B contains only critical actions, posterior-separability of the cost implies that any menu with more than I alternatives necessarily contains a non-critical option.¹⁵ The analyst exploits this by presenting the agent with elimination tasks $B \cup \{\mathbf{a}\}$ for various payoff vectors $\mathbf{a} \in \mathbb{R}^I \setminus B$. This ensures that the agent postpones learning and only eliminates non-critical alternatives.

Interpreting the data. Figure 5 depicts noiseless experimental data under base menu $B = \{\mathbf{b}^1, \mathbf{b}^2\}$ for an RI agent with prior $\boldsymbol{\pi}$. The point $\boldsymbol{\alpha}$ denotes the agent’s ignorance equivalent¹⁶ of the base menu. Payoff vectors $\times$ are rather unattractive; the agent dismisses them as noncritical in any menu $B \cup \{\times\}$. Payoff vectors $\diamond$ are

¹⁵Lemma A.2 in Appendix A.3 formalizes this claim.

¹⁶The concept of the ignorance equivalent is put forth by Müller-Itten et al. [2024]. In the main text, the authors focus on a more restrictive class of costs, but they include an existence and uniqueness proof under posterior separable costs in Online Appendix B.1.2.

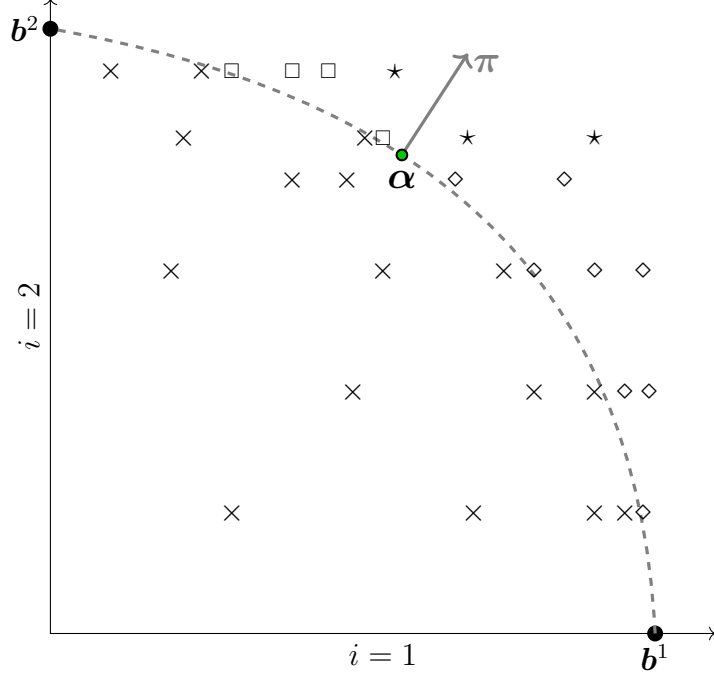


Figure 5: Cost and belief recovery from noiseless experimental data.

particularly attractive in state $i = 1$, prompting the agent to dismiss alternative $\mathbf{b}^1$ in menu $B \cup \{\diamond\}$. Payoff vectors $\square$ are particularly attractive in state $i = 2$, prompting the agent to dismiss $\mathbf{b}^2$ in $B \cup \{\square\}$. Finally, payoff vectors $\star$ are so attractive that they render both $\mathbf{b}^1$ and $\mathbf{b}^2$ noncritical. In the data, the analyst thus observes either $\mathbf{b}^1$ or $\mathbf{b}^2$ being dismissed.

The analyst tracks the payoffs and the agent's dismissal choice to recover the agent's unknown prior and cost: The shifted zero set boundary separates the payoff vectors that are dismissed from those that are retained. And along that boundary, the ignorance equivalent α is located where the agent switches from dismissing one base bet to another. Finally, the orthogonality vector at α identifies the agent's prior.

Data recording. The experiment can formally be described through the base menu B and a set of alternatives $D \subseteq \mathbb{R}^I \setminus B$. It starts with a series of elimination tasks with menus $B \cap \{\mathbf{a}\}$ for each $\mathbf{a} \in D$. At the end, one of the elimination tasks is revisited at random, and the agent selects one of the remaining options. To record the agent's behavior, the analyst assigns each payoff vector $\mathbf{a}$ to sets $(D^0, \dots, D^I)$ based on the agent's elimination choice: If she dismisses $\mathbf{b}^i \in B$, the payoff $\mathbf{a}$ is added to the corresponding D^i , and if she dismisses option $\mathbf{a}$ instead, the payoff vector is

added to D^0 .

Identification. For the formal result, I consider an idealized setting where the analyst observes countably many elimination tasks, such that $D = \bigcup_{i=0}^I D^i$ is dense in $\mathbb{R}^{\mathcal{I}}$. Compared with estimation techniques that rely on state-dependent stochastic choice data, this richness assumption is modest. After all, our analyst does not attempt to measure choice probabilities, he only records a single stochastic realization for each payoff vector in D .

Proposition 6. *Let $(D^0, \dots, D^I)$ describe the observed elimination choices in an experiment where $\bigcup_{i=0}^I D^i$ is dense in $\mathbb{R}^{\mathcal{I}}$. The following hold:*

- (a) *The ignorance equivalent α of base menu B is the unique element in $\bigcap_{i=0}^I \text{cl}(D^i)$.*
- (b) *The zero set is equal to $\text{cl}(D^0 - \{\alpha\})$.*

Proof. See Appendix A.3. □

Once the zero set is identified from the data, the prior and the cost potential can be deduced in the by now familiar way according to Proposition 2.

Risk aversion. Until this point, I have assumed that the agent is risk neutral, but the methods can be combined with standard experimental methods to accommodate other risk preferences. One way is to give stochastic rewards, the other is to separately elicit risk aversion and scale the payouts accordingly.

Perceived versus experienced costs. The elimination task asks the agent to choose between menus with varying incentives for learning, and even though the agent postpones the actual learning, her choice is guided by the cost that she associates with that learning. Since the agents anticipates these costs, her elimination choice is driven by her *perceived* rather than *experienced* information costs. This is much like in standard expected utility, where the agent's subjective belief dictates her choice under uncertainty, independent of whether her belief accurately captures the objective odds or not.

There are situations where the agent's perceptions is of primary interest, be it to better understand her worldview or to accurately predict her actions. Consider, for instance, the investor example from Müller-Itten et al. [2024]: An investor (she) can either directly invest a certain amount of money into one of many portfolios, or

can choose to retain an asset manager (he) who invests on her behalf. The asset manager may have a lower information cost and access to more portfolios, and he can commit to any state-dependent payout scheme. If both are risk-neutral, then the surplus-maximizing strategy for the asset manager is to offer the investor the ignorance equivalent dictated by the investor’s perceived learning costs. This ensures that she is willing to retain him unconditionally, and secures maximal rents for the manager. As such, the manager cares only about perceived costs; the investor’s actual learning costs are irrelevant for his payout design. In contrast, the investor’s self-directed and state-dependent direct investment choices would be dictated by her *experienced* learning costs that determine the extent to which she would actually learn about the market conditions before selecting a portfolio.

Caplin et al. [2020] develop an experimental approach to measure *experienced* learning costs. They repeatedly ask the agent to select bets from menus with varying incentives to learn, and thereby recover the stochastic choice data that results from their actual learning. The need to observe multiple contingencies of the ‘same’ choice problem motivates them to focus on simple production tasks – but in that setting, the two approaches would be comparable.

3 Experimental Results

(Coming Soon)

4 Conclusion

By representing an agent’s learning technology through a maximal set of zero welfare, I develop a novel experimental approach that is simple to implement and versatile. The protocol can be used to assess learning possibilities in real-world scenarios that are incompatible with existing elicitation techniques that seek to uncover the agent’s additive learning cost. At present, the paper is entirely theoretical and focused on the validity of the identification method, but a practical implementation is planned. The draft will be updated when those results are available.

References

- Ala Avoyan and Giorgia Romagnoli. Paying for inattention. *Economics Letters*, 226: 111114, 2023.
- Dimitri Bertsekas. *Convex optimization theory*, volume 1. Athena Scientific, 2009.
- Alexander W. Bloedel and Weijie Zhong. The cost of optimally acquired information. *Unpublished working paper*, 2020.
- Andrew Caplin, Dániel Csaba, John Leahy, and Oded Nov. Rational Inattention, Competitive Supply, and Psychometrics. *The Quarterly Journal of Economics*, 135 (3):1681–1724, 04 2020.
- Andrew Caplin, Mark Dean, and John Leahy. Rationally inattentive behavior: Characterizing and generalizing shannon entropy. *Journal of Political Economy*, 130(6): 1676–1715, 2022.
- Henrique de Oliveira, Tommaso Denti, Maximilian Mihm, and Kemal Ozbek. Rationally inattentive preferences and hidden information costs. *Theoretical Economics*, 12(2):621–654, 2017.
- Mark Dean and Nathaniel Neligh. Experimental tests of rational inattention. *Journal of Political Economy*, 131(12):3415–3461, 2023. doi: 10.1086/725174. URL <https://doi.org/10.1086/725174>.
- Tommaso Denti. Posterior separable cost of information. *American Economic Review*, 112(10):3215–59, October 2022. doi: 10.1257/aer.20211252. URL <https://www.aeaweb.org/articles?id=10.1257/aer.20211252>.
- Tommaso Denti, Massimo Marinacci, and Aldo Rustichini. Experimental cost of information. *American Economic Review*, 112(9):3106–23, September 2022. doi: 10.1257/aer.20210879. URL <https://www.aeaweb.org/articles?id=10.1257/aer.20210879>.
- Ambuj Dewan and Nathaniel Neligh. Estimating information cost functions in models of rational inattention. *Journal of Economic Theory*, 187:105011, 2020.
- Benjamin Hébert and Michael Woodford. Neighborhood-based information costs. *American Economic Review*, 111(10):3225–55, October 2021.

- Bartosz Maćkowiak and Mirko Wiederholt. Optimal sticky prices under rational inattention. *American Economic Review*, 99(3):769–803, 2009.
- Filip Matějka. Rationally inattentive seller: Sales and discrete pricing. *The Review of Economic Studies*, 83(3):1125–1155, 2016.
- Jianjun Miao and Hao Xing. Dynamic discrete choice under rational inattention. *Economic Theory*, pages 1–56, 2023.
- Jordi Mondria. Portfolio choice, attention allocation, and price comovement. *Journal of Economic Theory*, 145(5):1837–1864, 2010.
- Michèle Müller-Itten, Roc Armenter, and Zachary Stangebye. Rational inattention via ignorance equivalence. *Unpublished working paper*, 2024.
- Lin Peng. Learning with information capacity constraints. *Journal of Financial and Quantitative Analysis*, 40(2):307–329, 2005.
- Luciano Pomatto, Philipp Strack, and Omer Tamuz. The cost of information. *arXiv preprint arXiv:1812.04211*, 2018.
- R. Tyrrell Rockafellar. *Convex Analysis*. Princeton University Press, Princeton, New Jersey, 1970.
- Claude Elwood Shannon. A mathematical theory of communication. *The Bell system technical journal*, 27(3):379–423, 1948.
- Anthony F Shorrocks. The class of additively decomposable inequality measures. *Econometrica: Journal of the Econometric Society*, pages 613–625, 1980.
- Christopher A. Sims. Implications of rational inattention. *Journal of Monetary Economics*, 50(3):665–690, 2003.
- Luminita Stevens. Coarse pricing policies. *The Review of Economic Studies*, 87(1): 420–453, 07 2019. ISSN 0034-6527. doi: 10.1093/restud/rdz036. URL <https://doi.org/10.1093/restud/rdz036>.
- Elias Tsakas. Robust scoring rules. *Theoretical Economics*, 15(3):955–987, 2020.
- Stijn Van Nieuwerburgh and Laura Veldkamp. Information acquisition and under-diversification. *Review of Economic Studies*, 77(2):779–805, 2010.

A Additional Proofs and Results

A.1 Cost normalization

Proof of Lemma 1: I start with Claim (a), which normalizes the effective domain of the potential. To simplify notation, I define

$$C = \{\mathbf{w} \in \mathbb{R}^I \mid \boldsymbol{\pi} + t\mathbf{w} \in \text{dom}(\hat{\phi}^\pi) \text{ for some } t > 0\}$$

to capture the ‘feasible directions’ of belief movement. It is then possible to state the normalization equivalently as

$$\phi^\pi(\boldsymbol{\rho}) := \begin{cases} \hat{\phi}^\pi(\boldsymbol{\rho}) & \text{if } (\boldsymbol{\rho} - \boldsymbol{\pi}) \in C \cap (-C) \\ \infty & \text{otherwise.} \end{cases}$$

Since $\mathbf{0} \in C$, the finite value at the prior is maintained and ϕ^π satisfies Assumption A1. Note further that C is a convex cone since it is closed under linear combinations with nonnegative coefficients by convexity of $\text{dom}(\hat{\phi}^\pi)$ (which follows from applying Assumption A2 to $\hat{\phi}^\pi$). Consequently, the intersection $C \cap (-C)$ forms a linear subspace of $\mathbb{R}^n$.¹⁷ In particular, $C \cap (-C)$ is closed and convex.¹⁸ This allows us to express the epigraph as

$$\text{epi}(\phi^\pi) = \text{epi}(\hat{\phi}^\pi) \cap (C \cap (-C) \times \mathbb{R}).$$

As an intersection between two convex sets, $\text{epi}(\phi^\pi)$ is convex and hence satisfies Assumption A2. As an intersection between two closed sets, $\text{epi}(\phi^\pi)$ is closed and hence satisfies Assumption A3.

For Assumption A4, consider any $\boldsymbol{\rho} \in \text{dom}(\phi^\pi)$ that is not equal to the prior $\boldsymbol{\pi}$. By definition, $(\boldsymbol{\rho} - \boldsymbol{\pi}) \in -C$, and hence there exists $t > 0$ such that $\boldsymbol{\pi} + t(\boldsymbol{\pi} - \boldsymbol{\rho}) \in \text{dom}(\phi^\pi)$. By the Prolongation Lemma (see Footnote 10), Assumption A4 follows.

To show that ϕ^π maintains the cost function, it suffices to establish that $(\boldsymbol{\rho} - \boldsymbol{\pi}) \in C \cap (-C)$ with probability one for all $F \in \text{dom}(c^\pi)$, for then the two potentials maintain the effective domain of the cost function (on which they are identical anyways).

¹⁷Indeed, both C and $-C$ are closed under addition, and multiplying one of the $\mathbf{x} \in C \cap (-C)$ by any scalar can be recast as either multiplying $\mathbf{x}$ or $-\mathbf{x}$ (both of which are in $C \cap (-C)$) by a nonnegative scalar, so that it stays in the intersection cone.

¹⁸By definition, subspaces are closed under convex combinations. In $\mathbb{R}^n$, subspaces also contain all limits of their subsequences, so $C \cap (-C)$ is closed.

To do so, let $D = C \setminus (-C)$ denote the directions in which the two potentials ϕ^π and $\hat{\phi}^\pi$ differ. I first establish that D is convex. Indeed, let $\mathbf{v}, \mathbf{w} \in C \setminus (-C)$ and $t \in (0, 1)$. In particular, $-\mathbf{w} \in -C$. Since C is convex, $t\mathbf{v} + (1-t)\mathbf{w} \in C$. By contradiction, if $t\mathbf{v} + (1-t)\mathbf{w} \in -C$, then so is

$$\mathbf{v} = \frac{1}{t} \left[\underbrace{t\mathbf{v} + (1-t)\mathbf{w}}_{\in -C} + (1-t) \underbrace{(-\mathbf{w})}_{\in -C} \right].$$

After all, cones are closed under positive scalar multiplication and addition. In turn, this contradicts the assumption that $\mathbf{v} \in C \setminus (-C)$. Next, I establish that any distribution F with $\Pr(\{\boldsymbol{\rho} - \boldsymbol{\pi} \in D\} \mid \boldsymbol{\rho} \sim F) = p > 0$ is not in $\text{dom}(c^\pi)$. Indeed, $p > 0$ implies that the conditional expectation $\mathbf{v} = \frac{1}{1-p} \mathbb{E}[\boldsymbol{\rho} - \boldsymbol{\pi} \mid \boldsymbol{\rho} - \boldsymbol{\pi} \in D, \boldsymbol{\rho} \sim F]$ is well defined, and by convexity of the cone D , $\mathbf{v} \in D$. In particular, $\mathbf{v} \neq \mathbf{0}$. Since $F \in \Delta\Delta\mathcal{I} \mid \pi$, $p < 1$ and F is a garbling of the binary distribution that assigns probability p to posterior $\boldsymbol{\pi} + (1-p)\mathbf{v}$ and probability $1-p$ to posterior $\boldsymbol{\pi} - p\mathbf{v}$. Since the potential is convex, the cost of the garbling is bounded below by that of the binary distribution,

$$c^\pi(F) \geq p\hat{\phi}^\pi(\boldsymbol{\pi} + (1-p)\mathbf{v}) + (1-p)\hat{\phi}^\pi(\boldsymbol{\pi} - p\mathbf{v}).$$

Finally, $\mathbf{v} \in D$ and $D \cap (-C) = \emptyset$ implies that $-\mathbf{v} \notin C$ and hence $\hat{\phi}^\pi(\boldsymbol{\pi} - p\mathbf{v}) = \infty$. Since $c^\pi(F) = \infty$, this distribution indeed is not in the effective domain of c^π .

I now turn to Claim (b). As discussed on Page 9, Assumption A4 implies that $N_\pi(\hat{\phi}^\pi)$ contains a vector $(\mathbf{v}, -1)$. Since ϕ^π merely adds a linear function to $\hat{\phi}^\pi$, the transformation maintains the effective domain of the potential (Assumption A4) as well as convexity and closedness of the epigraph (Assumptions A2 and A3). At the prior, $\phi^\pi(\boldsymbol{\pi})$ evaluates to zero (Assumption A1), and

$$\begin{bmatrix} \mathbf{0} \\ -1 \end{bmatrix} \cdot \left(\begin{bmatrix} \boldsymbol{\rho} \\ s \end{bmatrix} - \begin{bmatrix} \boldsymbol{\pi} \\ \phi^\pi(\boldsymbol{\pi}) \end{bmatrix} \right) = \phi^\pi(\boldsymbol{\pi}) - s \leq 0 \quad \forall \begin{bmatrix} \boldsymbol{\rho} \\ s \end{bmatrix} \in \text{epi}(\phi^\pi)$$

by definition of the epigraph (Assumption A5).

To show that ϕ^π maintains the cost function, note that $\mathbb{E}[\phi^\pi(\boldsymbol{\rho}) \mid \boldsymbol{\rho} \sim F] - \phi^\pi(\boldsymbol{\pi})$ expands to

$$\mathbb{E}[\hat{\phi}^\pi(\boldsymbol{\rho}) \mid \boldsymbol{\rho} \sim F] + \underbrace{\mathbf{v} \cdot \mathbb{E}[\boldsymbol{\pi} - \boldsymbol{\rho} \mid \boldsymbol{\rho} \sim F]}_{=0} - \hat{\phi}^\pi(\boldsymbol{\pi}) \quad \forall F \in \Delta\Delta\mathcal{I} \mid \pi,$$

and hence is equal to c^π . □

Proof of Lemma 2: I start with the first claim. One direction is immediate: If there exists $(\mathbf{v}, -1) \in N_\pi(\phi^\pi)$ such that

$$\psi^\pi(\boldsymbol{\rho}) \equiv \phi^\pi(\boldsymbol{\rho}) + \mathbf{v} \cdot (\boldsymbol{\pi} - \boldsymbol{\rho}) \quad \forall \boldsymbol{\rho} \in \Delta\mathcal{I}, \quad (6)$$

then the two potentials represent the same cost,

$$\mathbb{E}[\phi^\pi(\boldsymbol{\rho}) \mid \boldsymbol{\rho} \sim F] \equiv \mathbb{E}[\psi^\pi(\boldsymbol{\rho}) \mid \boldsymbol{\rho} \sim F] \quad \forall F \in \Delta\Delta\mathcal{I} \mid_\pi. \quad (7)$$

After all, the affine term trivially cancels out in expectation.

For the other direction, suppose that Equation (7) holds. For any $\boldsymbol{\rho} \in \text{dom}(\phi^\pi)$, Assumption A4 imply the existence of some $\varepsilon > 0$ such that $\boldsymbol{\rho}^- = \boldsymbol{\pi} + \varepsilon(\boldsymbol{\pi} - \boldsymbol{\rho})$ also belongs to $\text{dom}(\phi^\pi)$. In particular, the binary strategy F with

$$F(\boldsymbol{\rho}) = \frac{\varepsilon}{1 + \varepsilon} \quad \text{and} \quad F(\boldsymbol{\rho}^-) = \frac{1}{1 + \varepsilon}$$

has finite cost. First, this implies that the posteriors also have a finite value under the other potential, and by symmetry of the argument, the effective domains $\text{dom}(\phi^\pi)$ and $\text{dom}(\psi^\pi)$ coincide. Second, Assumptions A2 and A3 imply the existence of a maximal $\bar{\varepsilon} : \Delta\mathcal{I} \rightarrow (0, \infty)$ such that $\boldsymbol{\pi} + \varepsilon(\boldsymbol{\pi} - \boldsymbol{\rho}) \in \text{dom}(\phi^\pi)$ if and only if $\varepsilon \leq \bar{\varepsilon}(\boldsymbol{\rho})$. Third, restating the cost implies a linear relationship between the difference in potentials on that segment of beliefs,

$$\begin{aligned} \frac{\varepsilon}{1 + \varepsilon} \phi^\pi(\boldsymbol{\rho}) + \frac{1}{1 + \varepsilon} \phi^\pi(\boldsymbol{\pi} + \varepsilon(\boldsymbol{\pi} - \boldsymbol{\rho})) &= \frac{\varepsilon}{1 + \varepsilon} \psi^\pi(\boldsymbol{\rho}) + \frac{1}{1 + \varepsilon} \psi^\pi(\boldsymbol{\pi} + \varepsilon(\boldsymbol{\pi} - \boldsymbol{\rho})) \\ \iff \psi^\pi(\boldsymbol{\pi} + \varepsilon(\boldsymbol{\pi} - \boldsymbol{\rho})) &= \phi^\pi(\boldsymbol{\pi} + \varepsilon(\boldsymbol{\pi} - \boldsymbol{\rho})) + \varepsilon[\phi^\pi(\boldsymbol{\rho}) - \psi^\pi(\boldsymbol{\rho})] \end{aligned} \quad (8)$$

for all $\varepsilon \in (0, \bar{\varepsilon}(\boldsymbol{\rho})]$.

Consider now a maximal list of affinely independent vectors $(\mathbf{b}^1, \dots, \mathbf{b}^K)$ such that $\boldsymbol{\pi} + \mathbf{b}^k \in \text{dom}(\phi^\pi)$. Since the effective domain is contained in $\Delta\mathcal{I}$, the length of the list is bounded above, $K \leq I$. Since the effective domains are identical, the equations

$$\phi^\pi(\boldsymbol{\pi} + \mathbf{b}^k) - \psi^\pi(\boldsymbol{\pi} + \mathbf{b}^k) = \mathbf{v} \cdot \mathbf{b}^k \quad \forall k = 1, \dots, K \quad (9)$$

are real-valued. As a linear system with $K \leq I$ constraints and an I -dimensional

unknown $\mathbf{v}$, it admits (at least) one solution $\mathbf{v} \in \mathbb{R}^{\mathcal{I}}$.

I now establish that Equation (6) holds for all $\boldsymbol{\rho} \in \Delta\mathcal{I}$:

- (i) If $\boldsymbol{\rho} \notin \text{dom}(\phi^\pi)$, equality of the potentials' effective domain implies $\psi^\pi(\boldsymbol{\rho}) = \infty$ and hence Equation (6) holds.
- (ii) If $\boldsymbol{\rho} = \boldsymbol{\pi} + \mathbf{b}^k$, Equation (6) is explicitly imposed as one of the constraints when choosing $\mathbf{v}$.
- (iii) If $\boldsymbol{\rho} = \boldsymbol{\pi} - \varepsilon \mathbf{b}^k \in \text{dom}(\phi^\pi)$ for some $\varepsilon \in (0, \bar{\varepsilon}(\boldsymbol{\pi} + \mathbf{b}^k)]$. In this case, Equation (8) evaluates to

$$\psi^\pi(\boldsymbol{\pi} - \varepsilon \mathbf{b}^k) = \phi^\pi(\boldsymbol{\pi} - \varepsilon \mathbf{b}^k) + \varepsilon[\phi^\pi(\boldsymbol{\pi} + \mathbf{b}^k) - \psi^\pi(\boldsymbol{\pi} + \mathbf{b}^k)] \stackrel{(9)}{=} \psi^\pi(\boldsymbol{\pi} - \varepsilon \mathbf{b}^k) + \mathbf{v} \cdot (\varepsilon \mathbf{b}^k),$$

establishing Equation (6).

- (iv) If $\boldsymbol{\rho}$ is any other belief in $\text{dom}(\phi^\pi)$, the vectors in the list $(\boldsymbol{\rho} - \boldsymbol{\pi}, \mathbf{b}^1, \dots, \mathbf{b}^K)$ are affinely dependent, meaning that there exist $\{m_k\}_{k=0}^K$ not all zero with

$$m_0(\boldsymbol{\rho} - \boldsymbol{\pi}) + \sum_{k=1}^K m_k \mathbf{b}^k = \mathbf{0} \quad \text{and} \quad \sum_{k=0}^K m_k = 0. \quad (10)$$

It is without loss of generality to assume that $m_0 > 0$, for the other vectors are affinely independent (ruling out $m_0 = 0$) and the equalities are maintained when all the coefficients are multiplied by (-1) .

Let $K^- = \{k \in \{1, \dots, K\} \mid m_k < 0\}$ and $K^+ = \{k \in \{1, \dots, K\} \mid m_k > 0\}$ denote the base vectors with negative and positive coefficients, respectively. Define $\underline{\varepsilon} = \min \{\bar{\varepsilon}(\boldsymbol{\pi} + \mathbf{b}^k) \mid k \in K^-\}$ and $\bar{m} = m_0 + \sum_{k \in K^+} m_k - \frac{1}{\underline{\varepsilon}} \sum_{k \in K^-} m_k$. As the minimum over finitely many positive terms, $\underline{\varepsilon} > 0$, and as the sum over more than one positive terms, $\bar{m} > 0$. Consider the learning strategy F with

$$F(\boldsymbol{\rho}) = \frac{m_0}{\bar{m}}, \quad F(\boldsymbol{\pi} + \mathbf{b}^k) = \frac{m_k}{\bar{m}} \forall k \in K^+ \quad \text{and} \quad F(\boldsymbol{\pi} - \underline{\varepsilon} \mathbf{b}^k) = \frac{-m_k}{\underline{\varepsilon} \bar{m}} \forall k \in K^-.$$

This strategy belongs to $\Delta\Delta\mathcal{I} \mid \boldsymbol{\pi}$ since all likelihoods are nonnegative, their sum equals one, and the expected posterior equals the prior,

$$\begin{aligned} \mathbb{E}[\boldsymbol{\sigma} - \boldsymbol{\pi} \mid \boldsymbol{\sigma} \sim F] &= \frac{m_0}{\bar{m}}(\boldsymbol{\rho} - \boldsymbol{\pi}) + \sum_{k \in K^+} \frac{m_k}{\bar{m}} \mathbf{b}^k + \sum_{k \in K^-} \frac{m_k}{\underline{\varepsilon} \bar{m}} \underline{\varepsilon} \mathbf{b}^k \\ &= \frac{1}{\bar{m}} \left[m_0(\boldsymbol{\rho} - \boldsymbol{\pi}) + \sum_{k=1}^K m_k \mathbf{b}^k \right] \stackrel{(10)}{=} \mathbf{0}. \end{aligned}$$

Since the two potentials represent the same cost,

$$\begin{aligned}
0 &= \mathbb{E} [\psi^\pi(\boldsymbol{\sigma}) - \phi^\pi(\boldsymbol{\sigma}) \mid \boldsymbol{\sigma} \sim F] \\
&= \frac{m_0}{\bar{m}}(\phi^\pi(\boldsymbol{\rho}) - \psi^\pi(\boldsymbol{\rho})) + \sum_{k \in K^+} \frac{m_k}{\bar{m}} \overbrace{(\phi^\pi(\boldsymbol{\pi} + \mathbf{b}^k) - \psi^\pi(\boldsymbol{\pi} + \mathbf{b}^k))}^{= \mathbf{v} \cdot \mathbf{b}^k \text{ by step (ii)}} \\
&\quad - \sum_{k \in K^-} \frac{m_k}{\underline{\varepsilon} \bar{m}} \overbrace{(\phi^\pi(\boldsymbol{\pi} - \underline{\varepsilon} \mathbf{b}^k) - \psi^\pi(\boldsymbol{\pi} - \underline{\varepsilon} \mathbf{b}^k))}^{= \mathbf{v} \cdot (-\underline{\varepsilon} \mathbf{b}^k) \text{ by step (iii)}} \\
&= \frac{m_0}{\bar{m}}(\phi^\pi(\boldsymbol{\rho}) - \psi^\pi(\boldsymbol{\rho})) + \mathbf{v} \cdot \underbrace{\left[\sum_{k \in K^+} \frac{m_k}{\bar{m}} \mathbf{b}^k + \sum_{k \in K^-} \frac{m_k}{\underline{\varepsilon} \bar{m}} \underline{\varepsilon} \mathbf{b}^k \right]}_{= -\frac{m_0}{\bar{m}}(\boldsymbol{\rho} - \boldsymbol{\pi}) \text{ by Equation (10)}} \\
&\Leftrightarrow \psi^\pi(\boldsymbol{\rho}) = \phi^\pi(\boldsymbol{\rho}) + \mathbf{v} \cdot (\boldsymbol{\pi} - \boldsymbol{\rho}),
\end{aligned}$$

establishing Equation (6).

To complete the first claim, it remains to show that $(\mathbf{v}, -1) \in N_\pi(\phi^\pi)$. As stated in the paragraph on normalization, Assumption A5 applied to ψ^π implies that $(\mathbf{0}, -1) \in N_\pi(\psi^\pi)$. Applied to ϕ^π , the assumption also implies that $\phi^\pi(\boldsymbol{\pi}) = 0$. From these, it follows readily that

$$\begin{bmatrix} \mathbf{v} \\ -1 \end{bmatrix} \cdot \left(\begin{bmatrix} \boldsymbol{\rho} \\ s \end{bmatrix} - \begin{bmatrix} \boldsymbol{\pi} \\ \phi^\pi(\boldsymbol{\pi}) \end{bmatrix} \right) \leq \mathbf{v} \cdot (\boldsymbol{\rho} - \boldsymbol{\pi}) - \phi^\pi(\boldsymbol{\rho}) \stackrel{(6)}{=} -\psi^\pi(\boldsymbol{\rho}) \stackrel{(A5)}{\leq} 0$$

for all $(\boldsymbol{\rho}, s) \in \text{epi}(\phi^\pi)$, and hence $(\mathbf{v}, -1) \in N_\pi(\phi^\pi)$.

To establish the equivalence between Claims (a) to (d), note first that $[(a) \Leftrightarrow (b)]$ follows directly from the previous argument. After all, ϕ^π and ψ^π coincide so long as all $(\mathbf{v}, -1) \in N_\pi(\phi^\pi)$ do not affect the value over the effective domain, i.e. $\mathbf{v} \cdot (\boldsymbol{\pi} - \boldsymbol{\rho}) = 0$ whenever $\boldsymbol{\rho} \in \text{dom}(\phi^\pi)$.

To establish the contrapositive to the implication $[(b) \Rightarrow (c)]$, I consider some $\boldsymbol{\rho} \in \text{dom}(\phi^\pi)$. By Assumption A4, the line segment from $\boldsymbol{\rho}$ to $\boldsymbol{\pi}$ can be extended beyond $\boldsymbol{\pi}$ without leaving the relative interior of $\text{dom}(\phi)$. Hence, there exists $\bar{\varepsilon} > 0$ such that the nonempty convex set

$$K = \left\{ \begin{bmatrix} \boldsymbol{\pi} \\ 0 \end{bmatrix} + \varepsilon \begin{bmatrix} \boldsymbol{\rho} - \boldsymbol{\pi} \\ \delta \end{bmatrix} \mid \varepsilon \in [-\bar{\varepsilon}, \bar{\varepsilon}] \right\}$$

is contained in $\text{dom}(\phi^\pi) \times \mathbb{R}$. If Claim (c) does not hold, then K has an empty intersection with the relative interior of the epigraph $\text{epi}(\phi^\pi)$. Hence, the Proper Separating Hyperplane Theorem¹⁹ implies the existence of a vector $(\mathbf{v}, w) \in \mathbb{R}^T \times \mathbb{R}$ such that

$$\begin{bmatrix} \mathbf{v} \\ w \end{bmatrix} \cdot \begin{bmatrix} \boldsymbol{\sigma} \\ s \end{bmatrix} \leq \begin{bmatrix} \mathbf{v} \\ w \end{bmatrix} \cdot \begin{bmatrix} \boldsymbol{\pi} + \varepsilon(\boldsymbol{\rho} - \boldsymbol{\pi}) \\ \varepsilon\delta \end{bmatrix} \quad \forall (\boldsymbol{\sigma}, s) \in \text{epi}(\phi^\pi), \forall \varepsilon \in [-\bar{\varepsilon}, \bar{\varepsilon}], \quad (11)$$

and the inequality being strict for at least one parameter combination $(\hat{\boldsymbol{\sigma}}, \hat{s}, \hat{\varepsilon})$.

Evaluating Equation (11) at $\boldsymbol{\sigma} = \boldsymbol{\pi}$, $s = 1$ and $\varepsilon = 0$ implies $w \leq 0$. Assume first by contradiction that $w = 0$, which allows us to restate Equation (11) as

$$\mathbf{v} \cdot (\boldsymbol{\sigma} - \boldsymbol{\pi}) \leq \varepsilon \mathbf{v} \cdot (\boldsymbol{\rho} - \boldsymbol{\pi}) \quad \forall \boldsymbol{\sigma} \in \text{dom}(\phi^\pi), \forall \varepsilon \in [-\bar{\varepsilon}, \bar{\varepsilon}]. \quad (11')$$

The strict inequality under $\hat{\boldsymbol{\sigma}} \in \text{dom}(\phi^\pi)$ and $\hat{\varepsilon} \in [-\bar{\varepsilon}, \bar{\varepsilon}]$ creates a contradiction with Equation (11') at other parameter values:

- If $\mathbf{v} \cdot (\hat{\boldsymbol{\sigma}} - \boldsymbol{\pi}) > 0$, then Equation (11') is violated at $\boldsymbol{\sigma} = \hat{\boldsymbol{\sigma}}$ and $\varepsilon = 0$.
- If $\mathbf{v} \cdot (\hat{\boldsymbol{\sigma}} - \boldsymbol{\pi}) < 0$, recall that Assumption A4 implies the existence of a $t > 0$ small enough such that $\boldsymbol{\sigma} = \boldsymbol{\pi} + t(\boldsymbol{\pi} - \hat{\boldsymbol{\sigma}}) \in \text{dom}(\phi^\pi)$, and Equation (11') is violated at this $\boldsymbol{\sigma}$ and $\varepsilon = 0$.
- If $\mathbf{v} \cdot (\hat{\boldsymbol{\sigma}} - \boldsymbol{\pi}) = 0$, then $\hat{\varepsilon} \mathbf{v} \cdot (\boldsymbol{\rho} - \boldsymbol{\pi}) > 0$, and Equation (11') is violated at $\hat{\boldsymbol{\sigma}}$ and $\varepsilon = -\hat{\varepsilon}$.

Hence, $w < 0$.

Evaluating Equation (11) for $\varepsilon = 0$, dividing both sides by $|w|$ and recalling that $\phi^\pi(\boldsymbol{\pi}) = 0$ by Assumption A5 establishes that $(\frac{1}{|w|}\mathbf{v}, -1)$ belongs to the normal cone $N_\pi(\phi^\pi)$. Yet, it is not possible that $\mathbf{v} \cdot (\boldsymbol{\pi} - \boldsymbol{\rho}) = 0$ for then evaluating Equation (11) at $\boldsymbol{\sigma} = \boldsymbol{\pi}$ and $s = 0$ would imply $0 \leq w\varepsilon\delta$ for both positive and negative values of ε . In other words, (b) does not hold either.

Assume now that (c) holds, and consider any $\mathbf{a} \perp \boldsymbol{\pi}$. Since unconditional implementation of the zero option is feasible in menu $\{\mathbf{0}, k\mathbf{a}\}$, the welfare $U(\{\mathbf{0}, k\mathbf{a}\} \mid \boldsymbol{\pi}, c^\pi)$ is always nonnegative. If it is positive for some $\hat{k} > 0$, then there exists a finite-cost learning strategy $F \in \Delta\Delta\mathcal{I} \mid_\pi$ that implements $\hat{k}\mathbf{a}$ with marginal probability $p \in (0, 1)$ at some posterior $\boldsymbol{\rho}^1 \in \text{dom}(\phi^\pi)$, and $\mathbf{0}$ otherwise. (Bayes rule pins down the opposite posterior, $\boldsymbol{\rho}^0 = \boldsymbol{\pi} + \frac{p}{1-p}(\boldsymbol{\pi} - \boldsymbol{\rho})$. Since the cost of the signal is finite,

¹⁹E.g., Proposition 1.5.6 in Bertsekas [2009].

$\boldsymbol{\rho}^0 \in \text{dom}(\phi^\pi)$.) Spelling out welfare yields

$$0 < U(\{\mathbf{0}, \hat{k}\mathbf{a}\} \mid \boldsymbol{\pi}, c^\pi) = \hat{k}p\boldsymbol{\rho}^1 \cdot \mathbf{a} - p\phi^\pi(\boldsymbol{\rho}^1) - (1-p)\phi^\pi(\boldsymbol{\rho}^0).$$

Since $\phi^\pi \geq 0$ by Assumption A5, the term $\boldsymbol{\rho}^1 \cdot \mathbf{a}$ is strictly positive. Consider now any other $k > 0$. The strategy that implements $k\mathbf{a}$ with marginal probability p at posterior $(1-t)\boldsymbol{\pi} + (1-t)\boldsymbol{\rho}^1$ and otherwise $\mathbf{0}$ at posterior $(1-t)\boldsymbol{\pi} + t\boldsymbol{\rho}^0$ achieves welfare

$$p \underbrace{((1-t)\boldsymbol{\pi} + t\boldsymbol{\rho}^1) \cdot (k\mathbf{a})}_{=kt\boldsymbol{\rho}^1 \cdot \mathbf{a} \text{ since } \mathbf{a} \perp \boldsymbol{\pi}} - p\phi^\pi(\boldsymbol{\pi} + t(\boldsymbol{\rho}^1 - \boldsymbol{\pi})) - (1-p)\phi^\pi(\boldsymbol{\pi} + t(\boldsymbol{\rho}^0 - \boldsymbol{\pi})).$$

By (c), there exists $t^1, t^0 > 0$ such that $\phi^\pi(\boldsymbol{\pi} + t^i(\boldsymbol{\rho}^i - \boldsymbol{\pi})) < t^i p k \boldsymbol{\rho}^1 \cdot \mathbf{a}$ for both i . By convexity of the potential, both inequalities also hold at $t = \min\{t^1, t^0\}$, and hence welfare is strictly larger than zero. The welfare under the optimal strategy is thus positive for all $k > 0$, establishing (d).

Finally, I establish the contrapositive of [(d) $\Rightarrow$ (b)]. To start, assume that there exists $(\mathbf{v}, -1) \in N_\pi(\phi^\pi)$ and $\boldsymbol{\rho} \in \text{dom}(\phi^\pi)$ such that $\mathbf{v} \cdot (\boldsymbol{\pi} - \boldsymbol{\rho}) \neq 0$. It is without loss of generality to assume that the scalar product is, in fact, positive – after all, Assumption A4 ensures that the line segment from $\boldsymbol{\rho}$ to $\boldsymbol{\pi}$ can be extended beyond $\boldsymbol{\pi}$ without leaving the effective domain, causing the sign to flip.

Let now $\mathbf{a} = \mathbf{v} - (\boldsymbol{\pi} \cdot \mathbf{v})\mathbf{1}$ with $\mathbf{a} \perp \boldsymbol{\pi}$. Consider any learning strategy that implements $\mathbf{a}$ with some marginal probability p at some posterior $\boldsymbol{\sigma}^1$ and $\mathbf{0}$ otherwise at some Bayes-plausible $\boldsymbol{\sigma}^0$. Net of learning costs, this strategy achieves welfare

$$\begin{aligned} & p[\boldsymbol{\sigma}^1 \cdot \mathbf{a} - \phi^\pi(\boldsymbol{\sigma}^1)] + (1-p)[\boldsymbol{\sigma}^0 \cdot \mathbf{0} - \phi^\pi(\boldsymbol{\sigma}^0)] \\ &= p[(\boldsymbol{\sigma}^1 - \boldsymbol{\pi}) \cdot \mathbf{v} - \phi^\pi(\boldsymbol{\sigma}^1)] + (1-p)[- \phi^\pi(\boldsymbol{\sigma}^0)] \end{aligned}$$

Since $(\mathbf{v}, -1)$ is in the normal cone, $\mathbf{v} \cdot (\boldsymbol{\sigma}^1 - \boldsymbol{\pi}) \leq \phi^\pi(\boldsymbol{\sigma}^1)$ regardless of the exact posterior $\boldsymbol{\sigma}^1$. And since $(\mathbf{0}, -1)$ is also in the normal cone by Assumption A5, $\mathbf{0} \cdot (\boldsymbol{\sigma}^0 - \boldsymbol{\pi}) = 0 \leq \phi^\pi(\boldsymbol{\sigma}^0)$. As the average between two nonpositive terms, the welfare of this generic strategy is nonpositive, and hence $W(\{\mathbf{0}, \mathbf{a}\} \mid \boldsymbol{\pi}, c^\pi) = 0$.

At the same time, Assumption A4 ensures that there exists a small enough $p \in (0, 1)$ such that there exists a strategy F that reaches $\boldsymbol{\rho}$ with marginal probability p , yet its other some posterior remains in the effective domain. Consequently, F has finite cost $c^\pi(F)$. As we scale up the difference between the two payoff vectors, the

welfare from this strategy grows without bounds,

$$\lim_{k \rightarrow \infty} \underbrace{p\sigma^1 \cdot k\mathbf{a}}_{\rightarrow \infty} - \underbrace{c^\pi(F)}_{\text{constant}} = \infty,$$

and becomes strictly positive eventually. This lower bound for $W(\{\mathbf{0}, k\mathbf{a}\} \mid \pi, c^\pi)$ implies a change in the sign of welfare, thereby negating (d). $\square$

A.2 Properties of the zero set

To avoid restating some common arguments in the subsequent proofs, I start by establish an auxiliary lemma.

Lemma A.1. *If ϕ^π admits a subgradient at $\rho \in \Delta\mathcal{I}$, then there exists $\mathbf{z}^\rho \in Z(\phi^\pi)$ with $\rho \cdot \mathbf{z}^\rho = \phi^\pi(\rho) = \sup_{\mathbf{z} \in Z(\phi^\pi)} \rho \cdot \mathbf{z}$.*

Proof. By definition, a subgradient of ϕ^π at ρ is a vector $\mathbf{v} \in \mathbb{R}^I$ such that

$$\phi^\pi(\sigma) - \phi^\pi(\rho) \geq \mathbf{v} \cdot (\sigma - \rho) \quad \forall \sigma \in \Delta\mathcal{I}$$

or equivalently, for $\mathbf{z}^\rho := \mathbf{v} + (\phi^\pi(\rho) - \mathbf{v} \cdot \rho)\mathbf{1}$,

$$\phi^\pi(\sigma) \geq \mathbf{z}^\rho \cdot \sigma \quad \forall \sigma \in \Delta\mathcal{I}.$$

By Equation (5), the payoff vector $\mathbf{z}^\rho$ thus belongs to the zero set.

Further, $\mathbf{z}^\rho \cdot \rho = \phi^\pi(\rho)$ and by Equation (5), this value represents an (achievable) upper bound for the supremum $\sup_{\mathbf{z} \in Z(\phi^\pi)} \rho \cdot \mathbf{z}$. $\square$

I now proceed with the proofs that are relegated from the main text.

Proof of Proposition 1: By Equation (5), the inequality $\rho \cdot \mathbf{a} \leq \phi^\pi(\rho)$ holds for any $\mathbf{a} \in Z(\phi^\pi)$ and any $\rho \in \Delta\mathcal{I}$. In turn, this implies that each value in the expectation (3) is nonpositive, regardless of the signal choice. In other words, the welfare from $Z(\phi^\pi)$ is at most zero – and since the normalized potential is nonnegative everywhere, $\mathbf{0} \in Z(\phi^\pi)$ ensures that welfare is exactly zero.

Conversely, fix any $\mathbf{a} \notin Z(\phi^\pi)$. By Equation (5), there exists $\rho \in \Delta\mathcal{I}$ such that $\phi^\pi(\rho) < \mathbf{a} \cdot \rho$. Since the prior π is in the relative interior of the effective domain $E = \text{dom}(\phi^\pi)$, the straight line segment from ρ to π can be extended beyond π without leaving E . Formally, there exists t small enough so that $\rho^- := \pi + t(\pi - \rho)$

also belongs to the relative interior of E . As a proper convex function, ϕ^π admits a subgradient at any point in the interior of its effective domain [Proposition 5.4.1 in Bertsekas, 2009], including ρ^- .

Consider now the signal that reaches posterior ρ with probability $\frac{t}{t+1}$ and ρ^- otherwise. (Such a signal exists since $\frac{t}{t+1}\rho + (1 - \frac{t}{t+1})\rho^- = \pi$.) By Equation (3), the agent's welfare from $Z(\phi^\pi) \cup \{a\}$ is bounded below by

$$\frac{t}{t+1} \underbrace{[\sup_{z \in Z(\phi^\pi)} \rho \cdot z - \phi^\pi(\rho)]}_{\geq \rho \cdot a > 0} + \left(1 - \frac{t}{t+1}\right) \underbrace{[\sup_{z \in Z(\phi^\pi)} \rho^- \cdot z - \phi^\pi(\rho^-)]}_{= 0 \text{ by Lemma A.1}} > 0,$$

and hence positive. $\square$

Proof of Proposition 2: To establish Claim (a), I first establish the contrapositive of the claim $[\pi \cdot z \leq 0 \ \forall z \in Z(\phi^\pi)]$. Indeed, consider any z with positive expected utility under the prior. By Equation (3), the agent's welfare from menu $Z(\phi^\pi) \cup \{z\}$ is bounded below by that achieved by the uninformative signal,

$$U(Z(\phi^\pi) \cup \{z\} \mid \pi, c^\pi) \geq \max_{a \in Z(\phi^\pi) \cup \{z\}} \pi \cdot a - \phi^\pi(\pi) \geq \pi \cdot z > 0.$$

It then follows from Proposition 1 that z cannot be contained in the zero set.

To show that the hyperplane $H = \{z \in \mathbb{R}^{\mathcal{I}} \mid \pi \cdot z = 0\}$ is balanced, I now extend the potential to all of $\mathbb{R}_{>0}^{\mathcal{I}}$ by homogeneity of degree one, $\bar{\phi}^\pi(t\rho) := t\phi^\pi(\rho)$ for any $\rho \in \Delta\mathcal{I}$. As it turns out, the intersection between H and $Z(\phi^\pi)$ is then exactly equal to the subdifferentials of the extension $\bar{\phi}^\pi$ at the prior.

Indeed, consider first any point $z \in H \cap Z(\phi^\pi)$. The following hold for any $t > 0$:

$$\begin{aligned} \bar{\phi}^\pi(t\rho) &= t\phi^\pi(\rho) \geq tz \cdot \rho \quad \forall \rho \in \Delta\mathcal{I} && \text{since } z \in Z(\phi^\pi) \\ -\bar{\phi}^\pi(\pi) &= -\phi^\pi(\pi) = -0 && \text{since } \phi^\pi \text{ is normalized} \\ -0 &= -z \cdot \pi && \text{since } z \in H. \end{aligned}$$

Summing these inequalities,

$$\bar{\phi}^\pi(t\rho) - \bar{\phi}^\pi(\pi) \geq z \cdot (t\rho - \pi) \quad \forall \rho \in \Delta\mathcal{I}, \forall t > 0$$

and hence z is a subgradient of $\bar{\phi}^\pi$ at the prior.

Conversely, suppose $\mathbf{z}$ is a subgradient of $\bar{\phi}^\pi$. This implies in particular that

$$\bar{\phi}^\pi(t\boldsymbol{\pi}) - \bar{\phi}^\pi(\boldsymbol{\pi}) = (t-1)\phi^\pi(\boldsymbol{\pi}) = 0 \geq \mathbf{z} \cdot (t\boldsymbol{\pi} - \boldsymbol{\pi}) = (t-1)\mathbf{z} \cdot \boldsymbol{\pi} \quad \forall t > 0,$$

which holds if and only if $\mathbf{z} \cdot \boldsymbol{\pi} = 0$, i.e. $\mathbf{z} \in H$. In addition,

$$\bar{\phi}^\pi(\boldsymbol{\rho}) - \bar{\phi}^\pi(\boldsymbol{\pi}) = \phi^\pi(\boldsymbol{\rho}) \geq \mathbf{z} \cdot (\boldsymbol{\rho} - \boldsymbol{\pi}) = \mathbf{z} \cdot \boldsymbol{\rho} - \underbrace{\mathbf{z} \cdot \boldsymbol{\pi}}_{=0} \quad \forall \boldsymbol{\rho} \in \Delta\mathcal{I},$$

and hence $\mathbf{z} \in Z(\phi^\pi)$.

In turn, since $\boldsymbol{\pi} \in \text{rel.int}(\text{dom}(\phi^\pi)) \subseteq \text{rel.int}(\text{dom}(\bar{\phi}^\pi))$, the set of subgradients of $\bar{\phi}^\pi$ at the prior can be written as the Minkowski sum of a linear subspace and a nonempty convex and compact set [Proposition 5.4.1 in Bertsekas, 2009].

As for Claim (b), fix any $\boldsymbol{\rho} \in \Delta\mathcal{I}$. The nonempty convex sets $\text{epi}(\phi^\pi)$ and $\{(\boldsymbol{\rho}, r)\}$ are both closed for any $0 \leq r < \phi^\pi(\boldsymbol{\rho})$. The latter set, a singleton, is bounded. By the definition of the epigraph in Equation (2), the intersection of the two sets is empty. Hence, the Strict Separation Theorem²⁰ implies that there exists a vector $(\mathbf{v}, t) \in \mathbb{R}^I \times \mathbb{R}$ and a scalar $\varepsilon > 0$ such that

$$(\mathbf{v}, t) \cdot (\boldsymbol{\sigma}, s) > (\mathbf{v}, t) \cdot (\boldsymbol{\rho}, r) + \varepsilon \quad \forall \boldsymbol{\sigma} \in \Delta\mathcal{I}, \forall s \geq \phi^\pi(\boldsymbol{\sigma}).$$

Evaluated at $\boldsymbol{\sigma} = \boldsymbol{\pi}$, the inequality implies $ts > \mathbf{v} \cdot (\boldsymbol{\rho} - \boldsymbol{\pi}) + tr + \varepsilon$ and hence ts is bounded below for any $s \geq \phi^\pi(\boldsymbol{\pi}) = 0$, no matter how large. This is possible only when $t \geq 0$.

After substituting $s = \phi^\pi(\boldsymbol{\sigma})$ when permissible, and rearranging terms,

$$t\phi^\pi(\boldsymbol{\sigma}) - tr - \varepsilon > \mathbf{v} \cdot \boldsymbol{\rho} - \mathbf{v} \cdot \boldsymbol{\sigma} \quad \forall \boldsymbol{\sigma} \in \Delta\mathcal{I} \text{ such that } \phi^\pi(\boldsymbol{\sigma}) < \infty. \quad (12)$$

Letting $\mathbf{z}^r := r\mathbf{1} + \frac{r}{tr+\varepsilon}[(\mathbf{v} \cdot \boldsymbol{\rho})\mathbf{1} - \mathbf{v}]$, we further obtain²¹

$$\begin{aligned} \mathbf{z}^r \cdot \boldsymbol{\sigma} &= r + \frac{r}{tr+\varepsilon}[\mathbf{v} \cdot \boldsymbol{\rho} - \mathbf{v} \cdot \boldsymbol{\sigma}] \\ &\stackrel{(12)}{\leq} r + \frac{r}{tr+\varepsilon}[t\phi^\pi(\boldsymbol{\sigma}) - tr - \varepsilon] = \frac{tr}{tr+\varepsilon}\phi^\pi(\boldsymbol{\sigma}) \leq \phi^\pi(\boldsymbol{\sigma}) \end{aligned}$$

for any $\boldsymbol{\sigma} \in \Delta\mathcal{I}$ with finite potential. And of course, the inequality $\mathbf{z}^r \cdot \boldsymbol{\sigma} \leq \phi^\pi(\boldsymbol{\sigma})$

²⁰See, for example, Proposition 1.5.3 in Bertsekas [2009].

²¹Recall that the normalized potential ϕ^π is nonnegative everywhere.

also holds when the right side is infinite. By Equation (5), $\mathbf{z}^r$ thus belongs to the zero set and

$$\mathbf{z}^r \cdot \boldsymbol{\rho} = r + \frac{r}{tr + \varepsilon} \underbrace{[\mathbf{v} \cdot \boldsymbol{\rho} - \mathbf{v} \cdot \boldsymbol{\rho}]}_{=0} = r$$

yields a lower bound for $\sup_{\mathbf{z} \in Z(\phi^\pi)} \boldsymbol{\rho} \cdot \mathbf{z}$.

Since the same conclusion is reached for any $r < \phi^\pi(\boldsymbol{\rho})$, the supremum is at least equal to $\phi^\pi(\boldsymbol{\rho})$. By the definition of the zero set in Equation (5), it is also no larger.

Finally, when ϕ^π admits a subgradient at $\boldsymbol{\rho}$, Lemma A.1 establishes that the supremum is attained. $\square$

Proof of Proposition 3: I start by establishing that $\boldsymbol{\pi}^Z$ and $\phi^Z(\boldsymbol{\rho})$ are well defined – that is, $\boldsymbol{\pi}^Z$ is a valid probability and ϕ^Z is a normalized potential that meets all the assumptions in Section 2.1.

- (i) Since $\mathbf{v}$ forms a supporting hyperplane, $\mathbf{v} \cdot \mathbf{z} \leq 0$ for all $\mathbf{z} \in Z$. Since $\mathbb{R}_{\leq 0}^{\mathcal{I}} \subseteq Z$, the negative base vector in any direction i is contained in Z , $-\mathbf{e}^i \in Z$ for any $i \in \mathcal{I}$. Together, the two observations imply that $\mathbf{v} \cdot (-\mathbf{e}^i) = -v_i \leq 0$, and hence $\mathbf{v} \geq \mathbf{0}$. Since $\mathbf{v}$ is nonzero by assumption, there exists at least one state j with $v_j > 0$, and hence $\mathbf{1} \cdot \mathbf{v} > 0$ and $\boldsymbol{\pi}^Z \in \Delta\mathcal{I}$.
- (ii) As the supremum of a nonempty set of linear functions, ϕ^Z is convex, lower semicontinuous and its image is contained in $\mathbb{R} \cup \{\infty\}$.

Since $\mathbf{0} \in Z$, the supremum $\sup_{\mathbf{z} \in Z} \boldsymbol{\rho} \cdot \mathbf{z}$ is nonnegative everywhere. At the prior, it is exactly zero since $\boldsymbol{\pi}^Z \in V$ and hence $\boldsymbol{\pi}^Z \cdot \mathbf{z} \leq 0$ for all $\mathbf{z} \in Z$. In other words, ϕ^Z satisfies the normalization assumptions.

Finally, it remains to show that the prior is in the relative interior of the effective domain. In line with [Proposition 1.3.3 in Bertsekas, 2009], we have to show that whenever $\boldsymbol{\rho} \in \text{dom}(\phi^\pi)$, the line segment from $\boldsymbol{\rho}$ to $\boldsymbol{\pi}$ can be prolonged beyond $\boldsymbol{\pi}$ without leaving the $\text{dom}(\phi^\pi)$, i.e. $\exists \varepsilon > 0$ such that $\boldsymbol{\pi} + \varepsilon(\boldsymbol{\pi} - \boldsymbol{\rho}) \in \text{dom}(\phi^\pi)$.

To do so, we use the fact that the hyperplane $H = \{\mathbf{z} \in \mathbb{R}^{\mathcal{I}} \mid \mathbf{z} \cdot \mathbf{v} = 0\}$ is balanced, allowing us to write $H \cap Z = L + K$ for a linear subspace L and a nonempty compact set K . Consider now any $\boldsymbol{\rho} \in \Delta\mathcal{I}$. If there exists $\mathbf{v} \in L$

such that $\mathbf{v} \cdot (\boldsymbol{\rho} - \boldsymbol{\pi}) \neq 0$, then for any $\mathbf{k} \in K$,

$$\phi^Z(\boldsymbol{\rho}) = \sup_{\mathbf{z} \in Z} \boldsymbol{\rho} \cdot \mathbf{z} \geq \sup_{\mathbf{z} \in Z \cap H} (\boldsymbol{\rho} - \boldsymbol{\pi}) \cdot \mathbf{z} + \underbrace{\sup_{r \in \mathbb{R}} r(\boldsymbol{\rho} - \boldsymbol{\pi}) \cdot \mathbf{v}}_{=\infty} + \underbrace{(\boldsymbol{\rho} - \boldsymbol{\pi}) \cdot \mathbf{k}}_{\text{bounded}} = \infty,$$

and there is nothing to show.

Otherwise, $\mathbf{v} \cdot (\boldsymbol{\rho} - \boldsymbol{\pi}) = 0 \ \forall \mathbf{v} \in L$ implies that $\sup_{\mathbf{z} \in H \cap Z} (\boldsymbol{\rho} - \boldsymbol{\pi}) \cdot \mathbf{z}$ is bounded above by some $M = \sup_{\mathbf{z} \in K} (\boldsymbol{\rho} - \boldsymbol{\pi}) \cdot \mathbf{z} < \infty$. Hence, the nonempty and convex sets Z and $H_M := \{\mathbf{h} \in H \mid (\boldsymbol{\rho} - \boldsymbol{\pi}) \cdot \mathbf{h} > M\}$ are disjoint for M large enough. By the separating hyperplane theorem, there exists a nonempty $\mathbf{v} \in \mathbb{R}^I$ and a real number $m \in \mathbb{R}$ such that

$$\mathbf{v} \cdot \mathbf{z} \leq m \leq \mathbf{v} \cdot \mathbf{h} \quad \forall \mathbf{z} \in Z, \forall \mathbf{h} \in H_M.$$

The right side is bounded below only in direction $(\boldsymbol{\rho} - \boldsymbol{\pi})$, it follows that $\mathbf{v} = k(\boldsymbol{\rho} - \boldsymbol{\pi})$ for some $k > 0$. In turn, this implies that the line segment from $\boldsymbol{\rho}$ to $\boldsymbol{\pi}$ can indeed be prolonged beyond $\boldsymbol{\pi}$ without leaving the effective domain of ϕ^Z ,

$$\phi^Z(\boldsymbol{\pi} + k(\boldsymbol{\pi} - \boldsymbol{\rho})) = \sup_{\mathbf{z} \in Z} (\boldsymbol{\pi} + k(\boldsymbol{\pi} - \boldsymbol{\rho})) \cdot \mathbf{z} \leq \underbrace{\sup_{\mathbf{z} \in Z} \boldsymbol{\pi} \cdot \mathbf{z}}_{=0} + \underbrace{\varepsilon \sup_{\mathbf{z} \in Z} k(\boldsymbol{\pi} - \boldsymbol{\rho}) \cdot \mathbf{z}}_{\leq m} < \infty.$$

Thus, the prior is indeed in the relative interior of the effective domain of ϕ^Z , and all normalization assumptions are satisfied.

Next, I show that $Z = Z(\phi^Z)$ in two steps.

- First, consider any $\mathbf{z}^0 \in Z$. By the definitions of ϕ^Z and the supremum,

$$\phi^Z(\boldsymbol{\rho}) = \sup_{\mathbf{z} \in Z} \boldsymbol{\rho} \cdot \mathbf{z} \geq \mathbf{z}^0 \cdot \boldsymbol{\rho} \quad \forall \boldsymbol{\rho} \in \Delta \mathcal{I},$$

and hence $\mathbf{z}^0 \in Z(\phi^Z)$ by Equation (5). And since $\mathbf{z}^0$ is arbitrary, $Z \subseteq Z(\phi^Z)$.

- Conversely, consider any $\mathbf{z}^0 \notin Z$. Since Z and $\{\mathbf{z}^0\}$ are disjoint and both nonempty, closed and convex. The latter set, a singleton, is bounded. Hence, the Strict Separation Theorem²² implies that there exists a vector $\mathbf{v} \in \mathbb{R}^I$ and

²²See, for example, Proposition 1.5.3 in Bertsekas [2009].

a scalar $r \in \mathbb{R}$ such that

$$\mathbf{v} \cdot \mathbf{z} < r < \mathbf{v} \cdot \mathbf{z}^0 \quad \forall \mathbf{z} \in Z. \quad (13)$$

By a similar argument as in point (i) above, the vector $\boldsymbol{\rho} = \frac{1}{1 \cdot \mathbf{v}} \mathbf{v}$ is a valid probability. Indeed, since $\mathbb{R}_{\leq 0}^Z \subseteq Z$, all negative multiples of the base vector in direction i are contained in Z . Inequality (13) thus implies in particular that $\mathbf{v} \cdot (-M\mathbf{e}^i) = -Mv_i$ is bounded above for any $M \geq 0$. This is possible only when $v_i \geq 0$, and the same argument applies for all states i . At the same time, $\mathbf{v} \neq 0$ or the left and right side in Equation (13) both evaluate to zero, violating the strict inequality. Thus, there also exists at least one state j with $v_j > 0$, and hence $1 \cdot \mathbf{v} > 0$. Taken together, $\boldsymbol{\rho}$ is well-defined, nonnegative, and its elements sum to one by definition. Applied to the definition of ϕ^Z , Equation (13) implies that

$$\phi^Z(\boldsymbol{\rho}) = \sup_{\mathbf{z} \in Z} \boldsymbol{\rho} \cdot \mathbf{z} = \frac{1}{1 \cdot \mathbf{v}} \sup_{\mathbf{z} \in Z} \mathbf{v} \cdot \mathbf{z} \leq \frac{r}{1 \cdot \mathbf{v}} < \frac{\mathbf{v} \cdot \mathbf{z}^0}{1 \cdot \mathbf{v}} = \boldsymbol{\rho} \cdot \mathbf{z}^0.$$

By the definition of $Z(\phi^Z)$ in Equation (5), $\mathbf{z}^0 \notin Z(\phi^Z)$ — and since $\mathbf{z}^0$ is arbitrary, $Z(\phi^Z) \subseteq Z$. $\square$

Proof of Proposition 4: Since both priors are on the relative interior of the effective domain of $\hat{\phi}$, Assumptions A1 and A4 are also satisfied. From Lemma 1(b) and Lemma 2, it follows that any normalized potential under prior $\boldsymbol{\pi}^i$ can be written as

$$\phi^i(\boldsymbol{\rho}) \equiv \hat{\phi}(\boldsymbol{\rho}) + \mathbf{v}^i \cdot (\boldsymbol{\pi}^i - \boldsymbol{\rho}) - \hat{\phi}(\boldsymbol{\pi}^i) \quad (14)$$

for some $(\mathbf{v}^i, -1) \in N_{\boldsymbol{\pi}^i}(\hat{\phi})$.

In particular, $\mathbf{z} \in Z(\phi^2)$ if and only if

$$\begin{aligned} \mathbf{z} \cdot \boldsymbol{\rho} &\leq \phi^2(\boldsymbol{\rho}) \stackrel{(14)}{=} \phi^1(\boldsymbol{\rho}) - \mathbf{v}^1 \cdot (\boldsymbol{\pi}^1 - \boldsymbol{\rho}) + \hat{\phi}(\boldsymbol{\pi}^1) + \mathbf{v}^2 \cdot (\boldsymbol{\pi}^2 - \boldsymbol{\rho}) - \hat{\phi}(\boldsymbol{\pi}^2) \\ \iff (\mathbf{z} + \mathbf{z}^0) \cdot \boldsymbol{\rho} &\leq \phi^1(\boldsymbol{\rho}) \iff \mathbf{z} + \mathbf{z}^0 \in Z(\phi^1) \end{aligned}$$

for $\mathbf{z}^0 = -\mathbf{v}^1 + \mathbf{v}^2 + (\mathbf{v}^1 \cdot \boldsymbol{\pi}^1 - \mathbf{v}^2 \cdot \boldsymbol{\pi}^2 - \hat{\phi}(\boldsymbol{\pi}^1) + \hat{\phi}(\boldsymbol{\pi}^2))\mathbf{1}$.

Moreover, $\mathbf{z}^0 \in Z(\phi^1)$ since

$$\phi^1(\boldsymbol{\rho}) - \mathbf{z}^0 \cdot \boldsymbol{\rho} = \underbrace{\phi^1(\boldsymbol{\rho}) - \mathbf{v}^1 \cdot (\boldsymbol{\pi}^1 - \boldsymbol{\rho}) + \hat{\phi}(\boldsymbol{\pi}^1)}_{=\hat{\phi}(\boldsymbol{\rho}) \text{ by (14)}} + \underbrace{\mathbf{v}^2 \cdot (\boldsymbol{\pi}^2 - \boldsymbol{\rho}) - \hat{\phi}(\boldsymbol{\pi}^2)}_{\geq -\hat{\phi}(\boldsymbol{\rho}) \text{ since } (\mathbf{v}^2, -1) \in N_{\boldsymbol{\pi}^2}(\hat{\phi})} \geq 0,$$

and for any $\mathbf{z} \in Z(\phi^1)$,

$$\mathbf{z} \cdot \boldsymbol{\pi}^2 \leq \phi^1(\boldsymbol{\pi}^2) \stackrel{(14)}{=} \hat{\phi}(\boldsymbol{\pi}^2) + \mathbf{v}^1 \cdot (\boldsymbol{\pi}^1 - \boldsymbol{\pi}^2) - \hat{\phi}(\boldsymbol{\pi}^1) = \mathbf{z}^0 \cdot \boldsymbol{\pi}^2.$$

In other words, $\mathbf{z}^0 \in \arg \max \{\boldsymbol{\pi}^2 \cdot \mathbf{z} \mid \mathbf{z} \in Z(\phi^1)\}$. $\square$

Proof of Proposition 5: Suppose first that $\mathbf{z}^0 \in Z(c^\pi)$. By menu-monotonicity of welfare, it follows trivially from Proposition 1 that $U(\{\mathbf{z}, \mathbf{0}\} \mid \boldsymbol{\pi}, c^\pi) \leq U(Z(c^\pi) \mid \boldsymbol{\pi}, c^\pi) = 0$.

Suppose now that $\mathbf{z}^0 \notin Z(c^\pi)$. For clarity, I break up the proof in distinct steps:

- Step 1: I identify a posterior $\boldsymbol{\rho} \in \Delta\mathcal{I}$ where the consumption gains from $\mathbf{z}^0$ exceed the increase in potential by more than $\delta > 0$.

Since the nonempty, closed, convex sets $Z(c^\pi)$ and $\{\mathbf{z}^0\}$ are disjoint and the latter is bounded, the Strict Separation Theorem [e.g. Proposition 1.5.3 in Bertsekas, 2009] implies that there exists a nonzero $\mathbf{v} \in \mathbb{R}^I$ and $k > 0$ with

$$\mathbf{v} \cdot \mathbf{z}^0 \geq \mathbf{v} \cdot \mathbf{z} + k \quad \forall \mathbf{z} \in Z(c^\pi).$$

By the now familiar argument, the fact that $\mathbb{R}_{\leq 0}^I \subseteq Z(c^\pi)$ implies that $\mathbf{v} \geq 0$, and since the vector is nonzero, $\mathbf{v} \cdot \mathbf{1} > 0$. In other words, $\boldsymbol{\rho} = \frac{1}{\mathbf{v} \cdot \mathbf{1}} \mathbf{v}$ is a valid probability with

$$\boldsymbol{\rho} \cdot \mathbf{z}^0 > \sup_{\mathbf{z} \in Z(c^\pi)} \boldsymbol{\rho} \cdot \mathbf{z} + \delta = \phi^\pi(\boldsymbol{\rho}) + \delta. \quad (15)$$

for any $\delta \in (0, \frac{1}{\mathbf{1} \cdot \mathbf{v}} k)$, with the last equality following from Proposition 2(b).

- Step 2: I find a compensating posterior $\boldsymbol{\sigma}$, on the other side of the prior, with a $\phi^\pi(\boldsymbol{\sigma})$ low enough relative to δ .

Equation (15) implies in particular that $\boldsymbol{\rho}$ is in the effective domain of ϕ^π . By Assumption A4, the line segment from $\boldsymbol{\rho}$ to $\boldsymbol{\pi}$ can be prolonged beyond $\boldsymbol{\pi}$ without leaving the effective domain, i.e. there exists $\bar{t} \in (0, 1)$ such that $\boldsymbol{\pi} + \bar{t}(\boldsymbol{\pi} - \boldsymbol{\rho}) \in \text{dom}(\phi^\pi)$. By Lemma 2(c), this in turn implies that there exists $\varepsilon \in (0, 1)$ such that $\phi^\pi(\boldsymbol{\pi} + \varepsilon \bar{t}(\boldsymbol{\pi} - \boldsymbol{\rho})) \leq \varepsilon \bar{t} \delta$. Let $t = \varepsilon \bar{t} \in (0, 1)$ and $\boldsymbol{\sigma} = \boldsymbol{\pi} + t(\boldsymbol{\pi} - \boldsymbol{\rho})$.

- Step 3: I conclude that $U(\{\mathbf{z}^0, \mathbf{0}\} \mid \boldsymbol{\pi}, c^\pi) > 0$.

Consider the learning strategy that implements $\mathbf{z}^0$ at posterior $\boldsymbol{\rho}$ with marginal probability $\frac{t}{t+1}$ and otherwise implements $\mathbf{0}$ at posterior. Net of learning costs, the utility from this strategy is equal to

$$\frac{t}{t+1} \underbrace{(\boldsymbol{\rho} \cdot \mathbf{z}^0 - \phi^\pi(\boldsymbol{\rho}))}_{>\delta \text{ by Equation (15)}} + \frac{1}{t+1} \underbrace{(0 - \phi^\pi(\boldsymbol{\pi} + t(\boldsymbol{\pi} - \boldsymbol{\rho})))}_{\geq -t\delta \text{ by Step 2}} > \frac{t}{t+1}(\delta - \delta) = 0.$$

Since this is a feasible strategy in menu $\{\mathbf{z}^0, \mathbf{0}\}$, it provides a positive lower bound for $U(\{\mathbf{z}^0, \mathbf{0}\} \mid \boldsymbol{\pi}, c^\pi)$.

This completes the proof. $\square$

A.3 Experimental Identification

Lemma A.2. *Any menu with $|A| > I$ options contains a non-critical alternative.*

Proof. This result is a direct consequence of Carathéodory's theorem. The RI optimization (3) determines the concave upper envelope of the (concave) net utilities $\boldsymbol{\rho} \mapsto \boldsymbol{\rho} \cdot \mathbf{a} - \phi^\pi(\boldsymbol{\rho})$ for posteriors $\boldsymbol{\rho}$ belonging to the $(|\mathcal{I}| - 1)$ -dimensional simplex $\Delta\mathcal{I}$. By Corollary 17.1.3 in Rockafellar [1970], this value can be expressed as a convex combination of posteriors that puts positive weight on at most I posteriors. In other words, there always exists an optimal signal that relies on at most $|\mathcal{I}|$ actions. Since $|B \cup \{\mathbf{a}\}| = |\mathcal{I}| + 1$, the result follows. $\square$

Lemma A.3. *Assumptions A7 and A8 imply that the optimal selection strategy in menu B relies on posteriors $\{\boldsymbol{\rho}^i\}_{i \in \mathcal{I}} \subseteq \Delta\mathcal{I}$ and associated marginals $\{m_i\}_{i \in \mathcal{I}}$ such that:*

- (a) $m_i > 0$ for all i , $\sum_{i \in \mathcal{I}} m_i = 1$ and $\sum_{i \in \mathcal{I}} m_i(\boldsymbol{\rho}^i - \boldsymbol{\pi}) = \mathbf{0}$.
- (b) For all i , $\mathbf{b}^i - \boldsymbol{\alpha} \in Z(c^\pi)$.
- (c) The vectors $\{\boldsymbol{\rho}^i - \boldsymbol{\pi}\}_{i \in \mathcal{I}}$ are affinely independent.
- (d) For each i , $\boldsymbol{\rho}^i$ is the unique posterior that is perpendicular to the zero set at $\mathbf{b}^i - \boldsymbol{\alpha}$,

$$\boldsymbol{\rho}^i \cdot \mathbf{z} \leq \boldsymbol{\rho}^i \cdot (\mathbf{b}^i - \boldsymbol{\alpha}) \quad \forall \mathbf{z} \in Z(c^\pi).$$

Proof. Assumption A7 implies that any optimal selection strategy over B relies on all I actions. Without loss of generality, one such strategy can be described by the

positive marginals m_i for each of I posteriors $\boldsymbol{\rho}^i$ with $\sum_{i \in \mathcal{I}} m_i \boldsymbol{\rho}^i = \boldsymbol{\pi}$, with alternative $\mathbf{b}^i$ being optimal under each posterior $\boldsymbol{\rho}^i$. This is Claim (a).

Since $\boldsymbol{\alpha}$ denotes the ignorance equivalent of the base menu,

$$\boldsymbol{\pi} \cdot \boldsymbol{\alpha} = U(B \cup \{\boldsymbol{\alpha}\} \mid \boldsymbol{\pi}, c^\pi) \stackrel{(4)}{\Longleftrightarrow} 0 = U((B - \{\boldsymbol{\alpha}\}) \cup \{\mathbf{0}\} \mid \boldsymbol{\pi}, c^\pi).$$

By menu monotonicity, the agent also achieves nonpositive welfare for any binary menu $\{\mathbf{b}^i - \boldsymbol{\alpha}, \mathbf{0}\}$, and so Claim (b) follows by Proposition 5.

To establish Claim (c), suppose that $\sum_{i \in \mathcal{I}} w_i (\boldsymbol{\rho}^i - \boldsymbol{\pi}) = \mathbf{0}$ and with $\sum_{i \in \mathcal{I}} w_i = 0$. Unless $w_i \equiv 0$, the supremum over the set $\{t \mid m_i + tw_i \geq 0 \quad \forall i\}$ is attained at a finite $\underline{t}$. By definition of the supremum, $m_i + \underline{t}w_i \geq 0$ also represent feasible marginals that achieve the same welfare. (After all, the weights sum to one and the expected posterior still equals the prior.) However, this strategy puts zero weight on one of the base vectors, contradicting Assumption A7. As a consequence, the vectors $\{\boldsymbol{\rho}^i - \boldsymbol{\pi}\}_{i \in \mathcal{I}}$ are affinely independent.

Claim (d) follows from Claim (b), for $B - \{\boldsymbol{\alpha}\} \subseteq Z(c^\pi)$ implies that the maximal welfare of zero is achieved only if the consumption gains exactly offset the cost potential at any posterior, i.e. $\boldsymbol{\rho}^i \cdot (\mathbf{b}^i - \boldsymbol{\alpha}) = \phi^\pi(\boldsymbol{\rho}^i)$. By the definition of the zero set, this in turn is weakly larger than $\boldsymbol{\rho}^i \cdot \mathbf{z}$ for all $\mathbf{z} \in Z(c^\pi)$. Moreover, $\boldsymbol{\pi} \cdot (\mathbf{b}^i - \boldsymbol{\alpha}) < 0$, for otherwise unconditional implementation of $\mathbf{b}^i$ would achieve maximal welfare, contradicting Assumption A7. By Assumption A8, the normal posterior is thus unique. $\square$

Proof of Proposition 6: Existence of an ignorance equivalent $\boldsymbol{\alpha}$ is established by Müller-Itten et al. [2024].²³ Suppose therefore that

$$U(B \mid \boldsymbol{\pi}, c^\pi) = U(\{\boldsymbol{\alpha}\} \mid \boldsymbol{\pi}, c^\pi) = U(B \cup \{\boldsymbol{\alpha}\} \mid \boldsymbol{\pi}, c^\pi). \quad (16)$$

By Assumption A7, unconditional implementation of any one alternatives in B yields strictly lower welfare, and so in particular, $\boldsymbol{\alpha} \notin B$. Equation (16) thus implies that none of the alternatives is critical in menu $B \cup \{\boldsymbol{\alpha}\}$, and hence $\boldsymbol{\alpha}$ itself may belong to any of the D^i depending on how the agent resolves her indifference. The same is not true when I instead consider any open ball O centered at $\boldsymbol{\alpha}$:

- Claim: There exists $\mathbf{a}^0 \in D \cap O$ such that $\mathbf{a}^0$ is the unique non-critical alternative in $B \cup \{\mathbf{a}^0\}$.

²³Posterior-separable costs are treated in their Online Appendix B.1.2.

Indeed, the intersection necessarily contains a vector $\mathbf{a}^0 \ll \boldsymbol{\alpha}$. By Assumption A7, dropping any of the base alternatives leads to a loss of welfare. At the same time, the welfare of menu $B \cup \{\mathbf{a}\}$ is trivially bounded above by that of $B \cup \{\boldsymbol{\alpha}\}$, which by definition of the ignorance equivalent is equal to that of B . In other words, dropping $\mathbf{a}^0$ does not lead to a welfare loss, dropping any of the other alternatives does. Hence $\mathbf{a}^0 \in D^0$.

- Claim: For each $i \in \mathcal{I}$, there exists $\mathbf{a}^i \in D \cap O$ such that $\mathbf{b}^i$ is the unique non-critical alternative in $B \cup \{\mathbf{a}^i\}$.

Let $\boldsymbol{\rho}^i \in \Delta\mathcal{I}$ be a posterior that is orthogonal to the zero set at $\mathbf{b}^i - \boldsymbol{\alpha}$. By Lemma A.3(d), this vector is unique. There exists $\mathbf{a}^i \in D \cap O$ such that $\mathbf{a}^i - \boldsymbol{\alpha}$ is extremal in direction $\boldsymbol{\pi} + \varepsilon(\boldsymbol{\rho}^i - \boldsymbol{\pi})$ for some $\varepsilon > 0$. If the agent drops alternative $\mathbf{b}^i$, she can maintain the same welfare as in B by using the feasible strategy $F \in \Delta\Delta\mathcal{I} \mid_{\boldsymbol{\pi}}$ with

$$F(\boldsymbol{\rho}^j) = \frac{\varepsilon m_j}{\varepsilon(1 - m_i) + m_i} \quad \forall j \neq i \quad \text{and} \quad F(\boldsymbol{\pi} + \varepsilon(\boldsymbol{\rho}^i - \boldsymbol{\pi})) = \frac{m_i}{\varepsilon(1 - m_i) + m_i}.$$

The argument works in reverse: If she could maintain welfare in menu $B \cup \mathbf{a}^i \setminus \{\mathbf{b}^j\}$ for some other $j \neq i$, then she could achieve the same welfare in menu $B \setminus \{\mathbf{b}^j\}$ by scaling down the marginals to compensate for the more extreme posterior $\boldsymbol{\rho}^j$ instead of $\boldsymbol{\pi} + \varepsilon(\boldsymbol{\rho}^i - \boldsymbol{\pi})$. This, in turn, would contradict Assumption A7.

Together, the two claims ensure that O has nonempty intersection with each D^i , irrespective of how the agent resolves her indifference. Since O is arbitrary, the ignorance equivalent $\boldsymbol{\alpha}$ is contained in $\bigcap_{i=0}^I \text{closure}(D^i)$.

Conversely, consider any $\mathbf{a} \in \bigcap_{i=0}^I \text{closure}(D^i)$. First, I show that $\mathbf{a} \notin B$. To this end, let δ be the minimal loss in welfare when one of the alternatives is removed from B . By Assumption A7, $\delta > 0$. For any payoff vector $\mathbf{b}^0$ that is located in an open ball centered at any of the $\mathbf{b}^i$ with radius δ , then its the achievable welfare in $B \cup \{\mathbf{b}^0\} \setminus \{\mathbf{b}^j\}$ is still smaller than that of B , for any $j \neq i$. After all, $\mathbf{b}^0$ brings less than δ additional consumption utility relative to $\mathbf{b}^i$, while the absence of $\mathbf{b}^j$ causes a loss of at least δ . In other words, there is an open ball around each $\mathbf{b}^i$ that is not contained in any of the D^j for $j \neq i$, and hence in particular, $\mathbf{a} \notin B$.

I now establish welfare comparisons that relate $\mathbf{a}$ to the ignorance equivalent of B . Indeed, by definition of the closure, each D^i contains a sequence $\{\mathbf{d}_k^i\}$ that

converges to $\mathbf{a}$ as $k \rightarrow \infty$. Since the limit is not part of the base menu, it is without loss of generality to assume that none of the elements in the sequence belong to B . Consequently, the menus $B \cup \{\mathbf{d}_k^i\}$ each contain $I + 1$ alternatives and Lemma A.2 ensures that the agent eliminates only noncritical alternatives. Hence, $\mathbf{d}_k^0 \in D^0$ implies that $U(B \mid \boldsymbol{\pi}, c^\pi) = U(B \cup \{\mathbf{d}_k^0\} \mid \boldsymbol{\pi}, c^\pi)$. By continuity of welfare, the same holds in the limit,

$$U(B \mid \boldsymbol{\pi}, c^\pi) = U(B \cup \{\mathbf{a}\} \mid \boldsymbol{\pi}, c^\pi). \quad (17)$$

Similarly, for any $i \in \mathcal{I}$, $\mathbf{d}_k^i \in D^i$ implies that $U(B \cup \{\mathbf{d}_k^i\} \setminus \{\mathbf{b}^i\} \mid \boldsymbol{\pi}, c^\pi) = U(B \cup \{\mathbf{d}_k^i\} \mid \boldsymbol{\pi}, c^\pi)$. Taking limits,

$$U(B \cup \{\mathbf{a}\} \setminus \{\mathbf{b}^i\} \mid \boldsymbol{\pi}, c^\pi) = U(B \cup \{\mathbf{a}\} \mid \boldsymbol{\pi}, c^\pi) \stackrel{(17)}{=} U(B \mid \boldsymbol{\pi}, c^\pi) \quad \forall i \in \mathcal{I}. \quad (18)$$

Consider now an optimal learning strategy over the base menu B , and let $\{\boldsymbol{\rho}^i\}_{i \in \mathcal{I}}$ denote the posteriors in its support and $m_i \in (0, 1)$ the associated marginal probability, with each $\boldsymbol{\rho}^i$ perpendicular to the zero set at the corresponding $\mathbf{b}^i - \boldsymbol{\alpha}$.

I first show that Equation (17) implies $\mathbf{a} - \boldsymbol{\alpha} \in Z(c^\pi)$. Suppose by contradiction that there exists some posterior $\boldsymbol{\sigma} \in \Delta\mathcal{I}$ with $\delta := \boldsymbol{\sigma} \cdot (\mathbf{a} - \boldsymbol{\alpha}) - \phi^\pi(\boldsymbol{\sigma}) > 0$. Since the vectors $\{\boldsymbol{\rho}^i - \boldsymbol{\pi}\}_{i \in \mathcal{I}}$ are affinely independent by Lemma A.3(c), one can write $\boldsymbol{\sigma} - \boldsymbol{\pi} = \sum_{i \in \mathcal{I}} w_i (\boldsymbol{\rho}^i - \boldsymbol{\pi})$ for some affine weights with $\sum_{i \in \mathcal{I}} w_i = 1$. Since all m_i are positive, there exists $\varepsilon > 0$ such that $m_i - \varepsilon w_i \in [0, 1]$ for all i . In other words, the strategy F with

$$F(\boldsymbol{\rho}^i) = m_i - \varepsilon w_i \quad \forall i \in \mathcal{I}, \quad F(\boldsymbol{\sigma}) = \varepsilon$$

is feasible. In menu $B \cup \{\mathbf{a}\}$, it enables the agent to choose $\mathbf{b}^i$ at posterior $\boldsymbol{\rho}^i$ and $\mathbf{a}$ at posterior $\boldsymbol{\sigma}$, achieving welfare

$$\begin{aligned} & \sum_{i \in \mathcal{I}} (m_i - \varepsilon w_i) (\boldsymbol{\rho}^i \cdot \mathbf{b}^i - \phi^\pi(\boldsymbol{\rho}^i)) + \varepsilon (\boldsymbol{\sigma} \cdot \mathbf{a} - \phi^\pi(\boldsymbol{\sigma})) \\ &= \sum_{i \in \mathcal{I}} (m_i - \varepsilon w_i) \underbrace{(\boldsymbol{\rho}^i \cdot (\mathbf{b}^i - \boldsymbol{\alpha}) - \phi^\pi(\boldsymbol{\rho}^i))}_{=0 \text{ by Lemma A.3(b) and (d)}} + \underbrace{\varepsilon (\boldsymbol{\sigma} \cdot (\mathbf{a} - \boldsymbol{\alpha}) - \phi^\pi(\boldsymbol{\sigma}))}_{=\delta} + \boldsymbol{\pi} \cdot \boldsymbol{\alpha} \\ &= \varepsilon \delta + U(B \mid \boldsymbol{\pi}, c^\pi) \end{aligned}$$

and thus contradicting Equation (17).

Next, I show that $\mathbf{a} - \boldsymbol{\alpha}$ is extremal to the zero set. Otherwise, any positive marginal weight on $\mathbf{a}$ at some posterior $\boldsymbol{\sigma}$ would lead to a net loss relative to uncon-

ditional implementation of α when accounting for the cost potential,

$$\sigma \cdot (a - \alpha) - \phi^\pi(\sigma) < 0,$$

and since $b^i - \alpha \in Z(c^\pi)$, this loss cannot be compensated at any other posteriors since $\rho \cdot (b^i - \alpha) - \phi^\pi(\rho) \leq 0$ for all $\rho \in \Delta\mathcal{I}$ and all i . Taken together, this means that the strategy that puts positive weight on a yields strictly lower welfare than $\pi \cdot \alpha = U(B \mid \pi, c^\pi)$, contradicting Equation (18). Even if the optimal strategy in $B \cup \{a\} \setminus \{b^i\}$ does not place any weight on a , its welfare is strictly lower than $U(B \mid \pi, c^\pi)$, for by Assumption A7, maximal welfare requires positive weight on *all* I alternatives in B , including b^i .

It remains to show that π is orthogonal to the zero set at $a - \alpha$. By contradiction, assume that $a - \alpha$ is not extremal in the π -direction, $\pi \cdot (a - \alpha) < 0$. By Assumption A8, the point then admits a unique normal vector σ . The same is true for any of the points $b^i - \alpha$ by Lemma A.3(d). Since I the vectors $\{\rho^i - \pi\}_{i \in \mathcal{I}}$ are affinely independent in the $I - 1$ dimensional space of vectors that are perpendicular to $\mathbf{1}$, it is possible to write $\sigma - \pi$ as a unique affine combination $\sum_{i \in \mathcal{I}} s_i(\rho^i - \pi)$ with $\sum_{i \in \mathcal{I}} s_i = 1$. In particular, there exists at least one $j \in \mathcal{I}$ with $s_j \leq m_j$. I now show that the optimal strategy in $B \cup \{a\} \setminus \{b^j\}$ cannot achieve the same welfare as the (hypothetical) strategy of implementing α always. Indeed, to avoid a net loss relative to this strategy after accounting for the cost potential, the optimal strategy would have to implement a at posterior σ and b^i at posterior ρ^j , for each $j \neq i$. By Bayes-plausibility, the marginal weights $\{w_i\} \subseteq [0, 1]$ with $\sum_{i=0}^I w_i = 1$ and $w_j = 0$ also have to satisfy

$$\sum_{i \neq j} w_i(\rho^i - \pi) + w_0(\sigma - \pi) = 0.$$

Subtracting $\sum_{i \in \mathcal{I}} m_i(\rho^i - \pi) = \mathbf{0}$ and substituting for $(\sigma - \pi)$, we obtain

$$\sum_{i \in I} (w_i + w_0 s_i - m_i)(\rho^i - \pi) = 0.$$

Since the coefficients sum to zero, affine independence of the vectors $\{\rho^i - \pi\}_{i \in \mathcal{I}}$ implies that they are all zero. In particular, coefficient j would have to satisfy $w_j + w_0 s_j - m_j = 0$. However, $w_j = 0$ and $w_0 \in [0, 1]$. Since $s_j \leq m_j$, the equality can hold only if $w_0 = 1$ — but if the agent puts full weight on a , it is implemented at posterior π , not σ .

In conclusion, I have shown that $\mathbf{a} - \boldsymbol{\alpha}$ is extremal in the zero set in the direction of the prior, hence

$$\boldsymbol{\pi} \cdot \boldsymbol{\alpha} = \boldsymbol{\pi} \cdot \mathbf{a} = U(\{\mathbf{a}\} \mid \boldsymbol{\pi}, c^\pi) = U(B \mid \boldsymbol{\pi}, c^\pi) \stackrel{(17)}{=} U(B \cup \{\mathbf{a}\} \mid \boldsymbol{\pi}, c^\pi),$$

implying that $\mathbf{a}$ itself is the ignorance equivalent of B . Claim (a) follows.

Claim (b) is simpler. Clearly, the agent is willing to dismiss any point $\mathbf{a}$ with $\mathbf{a} - \boldsymbol{\alpha} \in Z(c^\pi)$, for there is no posterior under which it brings a net gain relative to unconditional implementation of $\boldsymbol{\alpha}$. (And that unconditional implementation, in turn, yields welfare $U(B \mid \boldsymbol{\pi}, c^\pi)$.) If the vector is interior to the zero set, eliminating $\mathbf{a}$ is uniquely optimal by Assumption A7. In other words, $Z(c^\pi) \subseteq \text{cl}(D^0)$. For any $\mathbf{a} - \boldsymbol{\alpha} \notin Z(c^\pi)$, there exists a posterior $\boldsymbol{\sigma}$ with $\boldsymbol{\sigma} \cdot \mathbf{a}^0 > \phi^\pi(\boldsymbol{\sigma})$ for all $\mathbf{a}^0$ within an open ball O centered at $\mathbf{a}$. In this situation, dismissal of $\mathbf{a}^0$ is never optimal. Indeed, by dismissing $\mathbf{a}^0$, the agent secures welfare $U(B \mid \boldsymbol{\pi}, c^\pi) = \boldsymbol{\pi} \cdot \boldsymbol{\alpha}$. I now construct another dismissal strategy that does better. By affine independence (Lemma A.3(c)), it is possible to write $\boldsymbol{\sigma} - \boldsymbol{\pi}$ as an affine combination of the vectors $\{\boldsymbol{\rho}^i - \boldsymbol{\pi}\}_{i \in \mathcal{I}}$,

$$\boldsymbol{\sigma} - \boldsymbol{\pi} = \sum_{i \in \mathcal{I}} s_i (\boldsymbol{\rho}^i - \boldsymbol{\pi}) \quad \text{with} \quad \sum_{i \in \mathcal{I}} s_i = 1.$$

Let $\underline{t}$ denote the supremum of the set $\{t \mid m_i - ts_i \geq 0 \quad \forall i\}$. By Lemma A.3(a), this value is strictly positive. Consider now a selection strategy that relies on $F \in \Delta\Delta\mathcal{I} \mid_\pi$ with

$$F(\boldsymbol{\sigma}) = \underline{t}, \quad F(\boldsymbol{\rho}^i) = m_i - \underline{t}s_i \quad \forall i \in \mathcal{I}.$$

The agent selects $\mathbf{a}^0$ at $\boldsymbol{\sigma}$ and $\mathbf{b}^i$ at $\boldsymbol{\rho}^i$ and achieves strictly higher welfare

$$\underbrace{\underline{t}(\boldsymbol{\sigma} \cdot (\mathbf{a}^0 - \boldsymbol{\alpha}) - \phi^\pi(\boldsymbol{\sigma}))}_{>0} + \sum_{i \in \mathcal{I}} (m_i - \underline{t}s_i) \underbrace{(\boldsymbol{\rho}^i \cdot (\mathbf{b}^i - \boldsymbol{\alpha}) - \phi^\pi(\boldsymbol{\rho}^i))}_{=0 \text{ by Lemma A.3(d)}} + \boldsymbol{\pi} \cdot \boldsymbol{\alpha} > \boldsymbol{\pi} \cdot \boldsymbol{\alpha}.$$

However, by definition of $\underline{t}$, one of the base alternatives is never used, and so it is strictly better to eliminate that alternative instead of $\mathbf{a}^0$. This establishes Claim (b). $\square$

B Extensions

B.1 Zero set for select parametrizations

In this section, I derive closed form expressions for the zero set under select uniformly posterior separable cost functions found in the literature. I parametrize the upper boundary of the zero set via the function $\mathbf{z} : \Delta\mathcal{I} \times \Delta\mathcal{I} \rightarrow \mathbb{R}^{\mathcal{I}}$, in such a way that $Z(c^\pi) = \{\mathbf{z}(\boldsymbol{\rho} \mid \boldsymbol{\pi}) \mid \boldsymbol{\rho} \in \Delta\mathcal{I}\}$ and $\arg \max_{\mathbf{z}' \in Z(c^\pi)} \boldsymbol{\rho} \cdot \mathbf{z}' = \mathbf{z}(\boldsymbol{\rho} \mid \boldsymbol{\pi})$.

The key connection is the expression

$$\mathbf{z}(\boldsymbol{\rho} \mid \boldsymbol{\pi}) = \nabla\phi(\boldsymbol{\rho}) - \nabla\phi(\boldsymbol{\pi}) + (\phi(\boldsymbol{\rho}) - \phi(\boldsymbol{\pi}) + \nabla\phi(\boldsymbol{\pi}) \cdot \boldsymbol{\pi} - \nabla\phi(\boldsymbol{\rho}) \cdot \boldsymbol{\rho})\mathbf{1},$$

which follows from Equation (5) for differentiable cost potentials.

The Shorrocks [1980] generalized entropy cost used in Caplin et al. [2022], Dean and Neligh [2023], Hébert and Woodford [2021], Miao and Xing [2023] incorporates a curvature parameter $r \neq 1, 2$ in its cost potential

$$\phi^{\text{SE}}(\boldsymbol{\rho}) = \frac{1 - \sum_{i \in \mathcal{I}} \rho_i^{2-r}}{(r-1)(r-2)}.$$

In this case, Appendix B.1 evaluates to

$$z_k^{\text{SE}}(\boldsymbol{\rho} \mid \boldsymbol{\pi}) = \frac{\rho_k^{1-r} - \pi_k^{1-r}}{1-r} + \frac{\sum_i \pi_i^{2-r} - \sum_i \rho_i^{2-r}}{2-r}.$$

Mutual Information costs based on Shannon entropy are the most frequently used cost specification in applied RI work [e.g. Maćkowiak and Wiederholt, 2009, Matějka, 2016, Mondria, 2010, Peng, 2005, Stevens, 2019, Van Nieuwerburgh and Veldkamp, 2010]. Such costs correspond to the limit case of the Shorrocks's entropy when $r \rightarrow 1$, with potential $\phi^{\text{MI}}(\boldsymbol{\rho}) = \boldsymbol{\rho} \cdot \ln(\boldsymbol{\rho})$ and

$$\mathbf{z}^{\text{MI}}(\boldsymbol{\rho} \mid \boldsymbol{\pi}) = \ln(\boldsymbol{\rho}) - \ln(\boldsymbol{\pi}).$$

Total Information costs of Bloedel and Zhong [2020] incorporate parameters $\gamma_{ij} > 0$ that measure the ease with which the agent can discern between states i and j , with potential $\phi^{\text{TI}}(\boldsymbol{\rho}) = \sum_i \sum_j \rho_i \gamma_{ij} \ln\left(\frac{\rho_i}{\rho_j}\right)$ and

$$z_k^{\text{TI}}(\boldsymbol{\rho} \mid \boldsymbol{\pi}) = \sum_j \gamma_{kj} \left(\ln\left(\frac{\pi_j}{\pi_k}\right) - \ln\left(\frac{\rho_j}{\rho_k}\right) \right) + \sum_i \gamma_{ik} \left(\frac{\pi_i}{\pi_k} - \frac{\rho_i}{\rho_k} \right)$$

$$= \sum_j \left(\gamma_{kj} \ln \left(\frac{\pi_j}{\pi_k} \right) + \gamma_{jk} \frac{\pi_j}{\pi_k} \right) - \sum_j \left(\gamma_{kj} \ln \left(\frac{\rho_j}{\rho_k} \right) + \gamma_{jk} \frac{\rho_j}{\rho_k} \right).$$