Posterior-Separable Costs and Menu Preferences*

Henrique de Oliveira† — *FGV - São Paulo School of Economics*

Jeffrey Mensch‡ — *Hebrew University of Jerusalem*

November 14, 2025

Abstract

We consider an agent with a rationally inattentive preference over menus of acts, as in De Oliveira et al. (2017). We show that two axioms, *Independence of Irrelevant Alternatives* and *Ignorance Equivalence*, are necessary and sufficient for this agent to have a posterior-separable cost satisfying a mild smoothness condition, called joint-directional differentiability. Viewing the decision-maker's problem as a bayesian persuasion problem, we also show that these axioms are necessary and sufficient for solvability by a unique hyperplane. When the cost function remains invariant for different priors, we show that these axioms imply uniformly posterior separable costs that are differentiable.

1 Introduction

In many economic settings, a decision maker (DM) must acquire information before making a choice. This information may be costly, as it is difficult to acquire more precise information, while at the same time may allow for flexibility in the information chosen by focusing on different aspects of the problem. To this end, the rational inattention literature has focused to a large extent on *posterior separable* costs of information, where the cost of an experiment π is given by

$$c(\pi) = \int \psi(p) d\pi(p)$$

for some convex function. This representation allows for analyzing such problems in a tractable manner, using information design tools to solve for the problem, while guaranteeing that more information is always more costly in the Blackwell order (Blackwell, 1951). This paper aims to axiomatize preferences over menus that can be represented as coming from such costs.

There is already an existing literature on the representation of information costs based on individual preferences. Most directly connected is De Oliveira et al. (2017), which axiomatize preferences over menus that can be expressed as deriving from a DM who first acquires information, and then makes a choice from the menu that maximizes expected utility. This, in turn, is part of a more general approach of modeling

*This paper partially subsumes results from de Oliveira (2014), in which our two main axioms originally appear along with a partial characterization of the information costs that are consistent with them. We would like to thank Tommaso Denti for some helpful comments and suggestions, as well as seminar participants at SAET 2025.

†henrique.oliveira@fgv.br, https://henriquedeoliveira.com

‡jeffrey.mensch@mail.huji.ac.il, https://sites.google.com/site/jeffreyimensch/.

"costly contemplation" (Ergin and Sarver, 2010). Our work builds on these by providing axioms under which the representation in the former is posterior separable.

Separately, there is a literature on identifying whether a finite dataset is consistent with costly information acquisition through revealed preference. Caplin and Dean (2015) provide such a characterization for additively separable information costs, while Denti (2022) provides a characterization for posterior separable costs. Other papers, such as Caplin, Dean, and Leahy (2022) and Mensch and Malik (2024) characterize more specific forms of posterior-separable costs. These papers take as given a finite dataset, and check that for consistency with these respective theories via an acyclicality condition. By contrast, our paper takes preferences as given, and attempts to find a representation based on these.

Posterior-separable costs have found widespread application in the literature due to their ease of analysis. Under the hypothesis of expected utility, one can write the indirect utility as a convex function $\phi(p)$ of the posterior belief. As a result, the net objective of the DM at a given posterior is

$$N(p) := \phi(p) - \psi(p)$$

and so the DM's overall objective is linear in π. This allows for the use of *concavification* (Aumann and Maschler, 1995; Kamenica and Gentzkow, 2011) to solve for the optimal distribution, whereby the DM can achieve a value equal to the smallest concave function lying above N by choosing an appropriate distribution of posteriors. This appealing structure of the solution has led to its widespread use in applications in models with rationally inattentive agents. Some such work include Lipnowski, Mathevet, and Wei (2020), Yang (2020), Mensch (2022), Yoder (2022), Gleyze and Pernoud (2023), and Hébert and La'O (2023). In addition, more specific costs functions within this class have been developed, such as entropy costs (Sims, 2003; Matějka and McKay, 2015), neighborhood-based costs (Hébert and Woodford, 2021), and log-likelihood costs (Pomatto, Strack, and Tamuz, 2023). For a more extensive discussion of this class of functions, see Denti (2022).

Our result relies on two key axioms to tighten the representation of De Oliveira et al. (2017). The first axiom, *Independence of Irrelevant Alternatives*, states that if the DM is indifferent between two menus F, G as well as their intersection, then the DM is also indifferent to their union. The idea is that the preferences indicate that the options outside of the intersection are not useful, and so the DM does not benefit from the added flexibility. Thus, they remain without benefit when considering the union.

The second axiom, *Ignorance Equivalence*[1], states that for each menu, there is an act that the DM is indifferent to, including under union with the menu. This act provides an "ignorance equivalent" (Müller-Itten, Armenter, and Stangebye, 2023) to the menu. When this act is added to the menu, the agent is indifferent between acquiring information optimally for the menu, or not acquiring any information and simply choosing the act. Thus, the extra flexibility afforded by adding the act does not increase the menu's value. As Müller-Itten, Armenter, and Stangebye (2023) explain, the ignorance equivalent can be viewed as an analogue of a certainty equivalent for menus.

Together, these capture the idea that the optimal choice of the information by the DM, given that they are rationally inattentive, can be expressed via a unique hyperplane in the payoff-probability space. This hyperplane will be tangent to the payoff function at the optimal points of the information choice, and lie above the other points. Thus, the hyperplane is the unique optimal concavification of the DM's objective, given a posterior-separable objective. This "unique hyperplane" property holds if and only if the cost of information

[1] This axiom was called "Linearity" in de Oliveira (2014).

is both posterior separable and satisfies a property that we call "joint directional differentiability." We discuss the role of each axiom in the development of this conceptualization in Section 5.

The presence of a unique hyperplane is a common but underappreciated feature of the most commonly used posterior-separable cost functions, including the above-cited literature. Moreover, the set of such cost functions is dense with the set of posterior separable costs, and is satisfied whenever the cost function is differentiable (as is the case for almost all commonly used functions, such as entropy, log-likelihood, residual variance, and neighborhood-based costs). As we show in Section 6, when considering uniformly posterior-separable costs, our two axioms yield a differentiable cost representation.

This property is useful for solving rational inattention problems, as whenever the cost function is differentiable over all feasible distributions, one uses a first-order condition on the cost function to pin down the unique optimal hyperplane. The optimal hyperplane, in turn, indicates which posteriors can be chosen for each action from a given menu (see Caplin, Dean, and Leahy (2022), Lemma 1, which they refer to as the "Lagrangian lemma"). As the optimal choice of information is upper-hemicontinuous in the payoffs from the menu, the unique hyperplane property means that the Lagrange multiplier that defines the optimal choice of information will react continuously to perturbations of the menu. The well-behaved nature of information choice, given the unique hyperplane property, greatly simplifies the analysis of the DM's problem.

2 Model

Let Ω denote a finite set of states of the world and let X be a mixture space of consequences[2]. An *act* is a function $f : \Omega \to X$. A finite set of acts will be called a *menu* and denoted by F, G, H etc. The set of all acts is denoted by $\mathcal{F}$ and the set of all menus by $\mathbb{F}$. A single act f can also be seen as a singleton menu $\{f\}$; we usually omit the brackets if there is no chance for confusion.

Mixtures of acts are defined pointwise: given two acts f, g and a scalar $\alpha \in [0, 1]$, denote by $\alpha f + (1 - \alpha) g$ the act that in each state ω delivers the outcome $\alpha f(\omega) + (1 - \alpha) g(\omega)$. For $\alpha \in [0, 1]$, the mixture of two menus is defined as

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

We can interpret $\alpha F + (1 - \alpha) G$ as a lottery over what menu the agent faces.

Given an arbitrary set Z we let $\Delta(Z)$ denote the set of probability distributions over Z with finite support.

2.1 Rationally Inattentive Preferences

The primitive is a preference $\succsim$ defined over menus, which is interpreted according to the following timeline.

Thus, the agent chooses among menus while aware that they will be able to obtain information before finally choosing an act. We consider an agent who is rationally inattentive, that is, whose preferences can be represented by

$$V(\phi_F) = \max_{\pi \in \Pi(p_0)} \int \phi_F(p) \pi(dp) - c(\pi),$$

[2]For example, X could be the set of lotteries over a fixed set of prizes, or it could be a convex subset of some vector space.

where

$$\Pi(p_0) = \left\{ \pi \in \Delta(\Delta(\Omega)) \middle| \int p(\omega)\pi(dp) = p_0(\omega)\, \forall \omega \in \Omega \right\}$$

represents distributions over posterior beliefs consistent with possible finite information structures, $c : \Pi(p_0) \to \mathbb{R} \cup \{\infty\}$ is the cost of information[3], and

$$\phi_F(p) = \max_{f \in F} \sum_{\omega} u(f(\omega))p(\omega),$$

where $u : X \to \mathbb{R}$ is the utility function over consequences. We assume that the image of u is $\mathbb{R}$.

Note that, in the representation above, the consequence associated with an act in a given state only matters insofar as it affects the utility. Thus, we may consider as shorthand, instead of acts $f : \Omega \to X$, *utility acts* given by $u \circ f : \Omega \to \mathbb{R}$. We may also work with *utility menus*—finite sets of utility acts. For instance, when we refer to the menu $F = \{0\}$, we mean a utility menu that contains a single utility act $0 \in \mathbb{R}^\Omega$ or, equivalently, any menu that has a single act giving utility zero in every state.

The following result is proved in De Oliveira et al. (2017):

Proposition 1 (De Oliveira et al. (2017), Theorems 1 & 2). *Let $\succsim$ be a rationally inattentive preference. Then $\succsim$ has a representation (u, p_0, c) where the cost function $c : \Pi(p_0) \to \mathbb{R} \cup \{\infty\}$ is* canonical, *i.e. it satisfies*

Groundedness *$c(\delta_{p_0}) = 0$,*

Convexity *c is a convex function, and*

Blackwell-monotonicity *c is increasing in the Blackwell order.*

Moreover, this cost function can be recovered from the functional V by the formula

$$c(\pi) = \sup_{F \in \mathbb{F}} \int \phi_F(p)\pi(dp) - V(\phi_F).$$

Moreover, the following is a useful characterization of canonical costs, slightly modified from Denti, Marinacci, and Rustichini (2021), Lemma 6:

Proposition 2. *A cost function $c : \Pi(p_0) \to \mathbb{R} \cup \{\infty\}$ is canonical if and only if it can be written as*

$$c(\pi) = \sup_{\psi \in \Psi} \int \psi d\pi \tag{1}$$

where Ψ is a set of convex functions $\psi : \Delta(\Omega) \to \mathbb{R} \cup \{\infty\}$, such that

1. $\max_{\psi \in \Psi} \psi(p_0) = 0$, *and*

2. Ψ *is* minimal — *there is no $\hat{\Psi} \subset \Psi$ such that $\sup_{\psi \in \Psi} \int \psi d\pi = \sup_{\psi \in \hat{\Psi}} \int \psi d\pi$ and $|\hat{\Psi}| < |\Psi|$.*[4]

Proof. See Appendix A.1. □

[3]The value of ∞ is assigned to those distributions over posteriors that should never be acquired, representing impossible information. This could alternatively be modeled as a restriction on the domain of the cost function.

[4]We define $|\Psi| \in \mathbb{N} \cup \{\infty\}$ as the cardinality of the set Ψ, without making distinction between countable and uncountable infinities.

The main result in De Oliveira et al. (2017) is an axiomatic characterization of rationally inattentive preferences. The following properties will be useful for us:

1. If $F \subseteq G$ then $V(\phi_F) \leq V(\phi_G)$;

2. For any menu F and act h,
$$V(\phi_F + \phi_h) = V(\phi_F) + \phi_h(p_0).$$

2.2 Posterior-Separable Cost

Our goal in this paper is to understand a more specific class of costs of information, which we introduce now.

Definition 1. *The cost of information $c : \Pi(p_0) \to \mathbb{R} \cup \{\infty\}$ is said to be* posterior separable *if there exists an absolutely integrable function $\psi : \Delta(\Omega) \to \mathbb{R} \cup \{\infty\}$ such that, for all $\pi \in \Pi(p_0)$,*

$$c(\pi) = \int \psi(p)\pi(dp).$$

In this case, we say that c is represented by ψ, or that ψ is a representation of c, and call ψ a measure of uncertainty.

In the remainder of this subsection, we propose a class of functions ψ that have certain properties that are convenient to use and without loss of generality. Given a measure of uncertainty ψ, let

$$\operatorname{dom}\psi = \{p \in \Delta(\Omega) \mid \psi(p) < \infty\}.$$

Definition 2. *A measure of uncertainty $\psi : \Delta(\Omega) \to \mathbb{R} \cup \{\infty\}$ is* canonical *if:*

1. *ψ is convex;*

2. *$\psi(p_0) = 0$;*

3. *$\psi \geq 0$.*

4. *$p_0 \in \operatorname{ri}(\operatorname{dom}\psi)$;*

Before stating our formal result, we discuss some intuition for why these properties can be assumed without loss of generality. Convexity follows from Blackwell monotonicity of c. Property 2 follows from groundedness of c. Property 3 is a normalization that can be achieved by noting that adding an affine function to ψ that is zero at p_0 does not affect the cost of information. These first three properties are well known and often assumed whenever posterior-separable costs are used. To our knowledge, Property 4 is new. To see why it can be assumed without loss of generality, suppose that p_0 is not in the relative interior, as in Figure 1. There, p_0 lies in the vertical line that describes the left of the boundary of $\operatorname{dom}\psi$. Any π that puts positive probability on the right side of the line must also put positive probability on the left side of the line, where $\psi = \infty$, hence $c(\pi) = \infty$. Thus, if $c(\pi) < \infty$, it must be that the support of π is contained within the vertical line. This

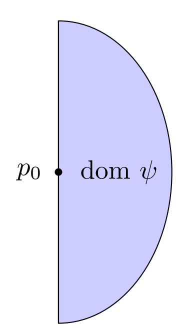


Figure 1: p_0 not in relative interior

means that nothing is lost by redefining the domain of ψ to be just the vertical line.

Below, we state the formal result, which also includes a sufficient condition for uniqueness.

PROPOSITION 3. *Let the information cost function c be canonical, as defined in Proposition 1. Suppose that c is posterior separable and represented by ψ. Then there exists a canonical measure of uncertainty $\tilde{\psi}$ such that c is represented by $\tilde{\psi}$. Moreover, if ψ is convex and differentiable in the directions of its domain at p_0, then $\tilde{\psi}$ is unique.*

Proof. See Appendix A.2.[5] □

From here onward, unless specified otherwise, we restrict our focus to posterior-separable costs that have a canonical measure of uncertainty.

2.3 Concavification

If the cost of information is posterior separable, we can write

$$V(\phi_F) = \max_{\pi \in \Pi(p_0)} \int N_F(p)\pi(dp)$$

where $N_F = \phi_F - \psi$ is the net utility. This parallels the objective in Bayesian persuasion, where the sender's objective is to find the optimal distribution with respect to the integrand. The optimum is found by taking the *concavification* of N_F, i.e.

$$V(\phi_F) = \text{cav}(N_F)(p_0) = \inf\{\zeta(p_0) : \zeta \geq N_F, \zeta \text{ concave}\}.$$

One can then use the techniques of finding the optimum from Bayesian persuasion, as found in Aumann and Maschler (1995) and Kamenica and Gentzkow (2011).

2.4 Dimension of Domain

As will become clear as we develop the main result, it will be useful to embed $\text{dom}(\psi)$ in a space that has the same dimension. To this end, let $\text{aff}(\text{dom}(\psi))$ be the affine hull of $\text{dom}(\psi)$ (the smallest affine set containing it) and let M be the dimension of this space. By Rockafellar (1970), Theorem 1.6, there exists a bijective affine transformation

$$T : \text{aff}(\text{dom}(\psi)) \to \mathbb{R}^M \tag{2}$$

For instance, in the case of full domain, where $\text{dom}(\psi) = \Delta(\Omega) \subset \mathbb{R}^\Omega$, one can define such a T as mapping to $\mathbb{R}^{|\Omega|-1}$ by setting $T_i(p) = p(\omega_i)$ for $i \in \{1, ..., |\Omega| - 1\}$, implicitly determining the probability of the remaining state $\omega_{|\Omega|}$ as $1 - \sum_{i<|\Omega|} p(\omega_i)$.

Given any $\psi : \Delta(\Omega) \to \mathbb{R} \cup \{\infty\}$, we define $T^*\psi : \mathbb{R}^M \to \mathbb{R} \cup \{\infty\}$ by

$$T^*\psi(y) = \begin{cases} \psi(T^{-1}(y)), & y \in T(\text{dom}(\psi)) \\ \infty, & \text{otherwise} \end{cases} \tag{3}$$

[5]The proof uses a construction shown in section 2.4, so the reader is advised to read that section before reading the proof.

We call $T^*\psi$ the translation of ψ to $\mathbb{R}^M$. Notice that, since T is a linear transformation, the properties of a canonical ψ are slightly modified by $T^*\psi$ as follows:

1. $T^*\psi$ is convex;
2. $T(p_0) \in \text{int}(T(\text{dom}(\psi))$;
3. $T^*\psi(T(p_0)) = 0$;
4. $T^*\psi \geq 0$.

Notice, in particular, that since $\dim(T(\text{dom}(\psi))) = M$, $T(p_0)$ is now in the *interior* of $T(\text{dom}(\psi))$, not just the relative interior. To economize on notation, we write $\bar{\psi} \coloneqq T^*\psi$, $\bar{p} \coloneqq T(p)$, and $Y \coloneqq T(\text{dom}(\psi))$.

2.5 Unique Hyperplane Property

In this section, we present an alternative description of concavification, equivalent to that presented in Section 2.3. This relies on the supporting hyperplane of the concave function $\text{cav}(N_F)$, so that all values of the function lie below this hyperplane.[6] We use this representation of the concavification to develop key properties of our cost function representation.

Let T be as in (2). For any hyperplane $\mathcal{H} \in \mathbb{R}^{M+1}$, let λ be its normal vector. Define $N_F^* : \mathbb{R}^M \to \mathbb{R}$ by

$$N_F^*(y) = \begin{cases} N_F(T^{-1}(y)), & y \in Y \\ -\infty, & \text{otherwise} \end{cases} \tag{4}$$

An equivalent formula for the concavification for a given prior p_0 is[7]

$$\text{cav}(N_F^*)(\bar{p}_0) = \min_{\substack{\lambda \in \mathbb{R}^M \\ k \in \mathbb{R}}} \{\lambda \cdot \bar{p}_0 + k : \lambda \cdot y + k \geq N_F^*(y), \forall y \in Y\}. \tag{5}$$

A pair (λ, k) is a solution to the minimization problem above if $\lambda \cdot y + k \geq N_F^*(y)$ for all $y \in Y$ and $\lambda \cdot \bar{p}_0 + k = \text{cav}(N_F^*)(\bar{p}_0) = V(\phi_F)$. Thus, we can solve for k in this expression and denote the set of solutions to (5) by

$$\Lambda_F = \left\{\lambda \in \mathbb{R}^M | \lambda \cdot (y - \bar{p}_0) + V(\phi_F) \geqslant N_F^*(y), \forall y \in Y\right\}$$

By Rockafellar (1970), Theorem 23.2, Λ_F is closed, convex, and non-empty. Geometrically, the elements of Λ_F are the normal vectors of hyperplanes that are tangent to the graph of N_F^* (see fig. 2). When N_F^* is well-behaved (in a sense that will be made precise shortly), this set is a singleton, motivating the following definition:

Definition 3. *ψ satisfies the* unique hyperplane property (UHP) *if, for all menus F, Λ_F is a singleton.*

2.6 Joint-directional Differentiability

As will become clear later, the unique hyperplane property is related to a notion of differentiability that we now discuss.

[6] This is analogous to the "Lagrangian lemma" of Caplin, Dean, and Leahy (2022).

[7] This follows from Rockafellar (1970), Theorem 18.8, which states that any closed convex set in a Euclidean space is the intersection of the closed half-spaces tangent to it. Since $\text{epi}(-\text{cav}(N_F))$ is closed and convex, there exist such tangent hyperplanes as described in (5) for $p = p_0$.

Definition 4. *Let* $\bar{\psi} : Y \to \mathbb{R} \cup \{\infty\}$ *be as defined in Section 2.4. The* subdifferential *of* $\bar{\psi}$ *at* $y \in Y$ *is*

$$\partial\bar{\psi}(y) = \{\lambda \in \mathbb{R}^M : \lambda \cdot (q - y) \le \bar{\psi}(q) - \bar{\psi}(y), \forall q \in Y\}.$$

A convex function is differentiable if and only if its subdifferential is a singleton (Rockafellar (1970), Theorem 25.1). To compare the subdifferentials at various points when they are not singletons, we introduce the following property of ψ.

Definition 5. *The function* $\bar{\psi}$ *is* non-differentiable in the direction $\delta \in \mathbb{R}^M \setminus \{0\}$ at $\{\bar{p}_i\}_{i=1}^K$ *if there exist* $\lambda_i \in \partial\bar{\psi}(\bar{p}_i)$ *such that:*

1. $\lambda_i + \delta \in \partial\bar{\psi}(\bar{p}_i), \forall i \in \{1, ..., K\}$ *and*
2. $\delta \cdot (\bar{p}_i - \bar{p}_0) = 0, \forall i \in \{1, ..., K\}$.

We say that $\bar{\psi}$ *is* non-differentiable in the same direction (NDISD) at $\{\bar{p}_i\}_{i=1}^K$ *if there exists a* $\delta \in \mathbb{R}^M \setminus \{0\}$ *such that* $\bar{\psi}$ *is non-differentiable in the direction* δ *at* $\{\bar{p}_i\}_{i=1}^K$.

Geometrically, we can think of an element of the subdifferential as the slope of a hyperplane that is tangent to the graph of $\bar{\psi}$ at the point $\bar{p}_i$. When there is more than one such tangent hyperplane, it means that the function $\bar{\psi}$ has a kink at that point, and we can think of the difference between the two slopes (δ) as a direction of that kink, in the sense that one can wobble the hyperplane by adding δ and remain tangent. The condition above states that at all the points $\{\bar{p}_i\}_{i=1}^K$ there is a kink and that they all share a wobbling direction δ. Thus, the direction δ is orthogonal to the convex hull of $\{\bar{p}_i - \bar{p}_0\}_{i=1}^K$.

This definition will be useful in constructing menus that take advantage of the directions of non-differentiability to violate our axioms. In particular, it enables us to define cost functions that are "sufficiently smooth."

Definition 6. *Given prior* p_0*, the function* ψ *satisfies* joint-directional differentiability (JDD) *if there do not exist* $\delta \in \mathbb{R}^M \setminus \{0\}$ *and points* $\{p_i\}_{i=1}^K$*, with* $p_0 \in co(p_1, \ldots, p_K)$*, such that* $\bar{\psi}$ *is non-differentiable in the direction* δ *at* $\{\bar{p}_i\}_{i=1}^K$.

Remark: Whether ψ satisfies JDD does not depend on the choice of T. Suppose $\hat{T}$ is another choice. Because T and $\hat{T}$ have the same dimension in the domain and codomain, there must be an invertible linear transformation $A : \mathbb{R}^M \to \mathbb{R}^M$ such that $\hat{T} = A \circ T$. Letting $\mathbf{A}$ be the matrix associated with A, if $\lambda \in \partial\bar{\psi}(x)$, then $(\mathbf{A}^{-1})^\intercal \lambda \in \partial(A \circ T)^*\psi$, where $(\mathbf{A}^{-1})^\intercal$ is the transpose of the inverse of $\mathbf{A}$.

3 Main Theorem

3.1 Axioms

Our main result relies on the following axioms.

Independence of Irrelevant Alternatives (IIA). *If* $F \sim F \cap G \sim G$ *then* $F \sim F \cup G$.

For any menus F and G, we always have $F \cup G \succsim F, G \succsim F \cap G$, because flexibility is never harmful. When $F \sim F \cap G$, we may say that the elements of F that do not belong to $F \cap G$ are irrelevant: the DM can achieve the same payoff even by ignoring these additional options. The axiom states that these irrelevant options remain irrelevant when combined with other irrelevant options.

Ignorance Equivalence (IE). *For every menu F, there exists an act h such that $h \sim F \sim F \cup h$.*

When faced with a singleton menu h, it is always optimal for the agent to acquire no information. The act h in this axiom is just as good as F, yet it adds irrelevant flexibility to F. Thus, the act h can be thought of as an ignorance equivalent—a version of the menu F that requires no information to be acquired.

First appearing in de Oliveira (2014), this property also appears in Müller-Itten, Armenter, and Stangebye (2023), who coined the term "ignorance equivalent" to refer to h. The term comes from an analogy with the role of a certainty equivalent in the context of decisions under risk. Indeed, in both contexts, a risk-neutral principal can extract the most possible surplus by offering the equivalents: by offering the ignorance-equivalent in the former, and by offering the certainty equivalent in the latter (i.e. fully insuring). The ignorance equivalent also serves as a tool to see which actions might ever be chosen in an expanded menu: when comparing F and $F \cup \{g\}$, it is sufficient to know that h, the ignorance equivalent, dominates g (i.e. is better in every state) in order to conclude that g will never be chosen from $F \cup \{g\}$, and therefore $F \sim F \cup \{g\}$.

3.2 Main Theorem

We can now state our main result:

Theorem 1. *The following statements are equivalent:*

1. *$\succsim$ is a rationally inattentive preference satisfying IIA and IE;*
2. *$\succsim$ has a posterior-separable representation with a measure of uncertainty ψ satisfying joint-directional differentiability.*
3. *$\succsim$ has a posterior-separable representation with a measure of uncertainty ψ satisfying the unique hyperplane property.*

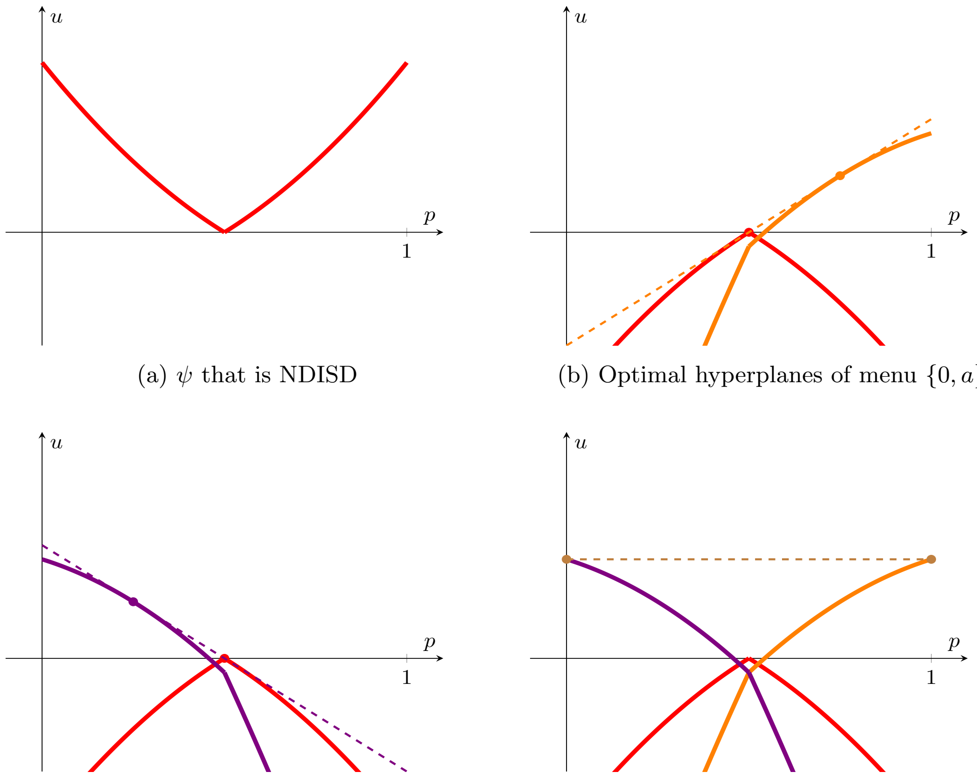
3.3 Example

To see how the three statements in Theorem 1 are related, we present the following example. In particular, we highlight the role that the unique hyperplane property has in our analysis, showing that, as the unique hyperplane property is violated, this leads to a violation of IIA.

Consider a setup with two states, in which the prior $p_0 = 0.5$ and the cost of information is given (Figure 1a) by

$$\psi(p) = |p - 0.5| + (p - 0.5)^2$$

As ψ is non-differentiable at the prior, it violates joint directional differentiability. Indeed, with the menu $F = \{0\}$, there are multiple optimal hyperplanes, all of which yield $V(F) = 0$. Now suppose we consider two new actions $\{a, b\}$, with respective payoffs $u(a, p) = 2.5p - 1.3125$ and $u(b, p) = -2.5p + 1.1875$. One easily verifies that the DM is indifferent between the menus $\{0\}$, $\{0, a\}$, and $\{0, b\}$, as seen by the fact that all three generate respective optimal hyperplanes λ such that the value of λ at $p_0 = 0.5$ is 0 (Figures 1b and 1c). However, if we take the union $\{0, a, b\}$, the optimal hyperplane changes: it now becomes optimal to choose posteriors $p \in \{0, 1\}$ (Figure 1d), yielding a value of the menu of 0.375. As a and b are therefore only relevant when we take the union of the menus, but not when we add them individually to 0, IIA is not satisfied.



(a) ψ that is NDISD

(b) Optimal hyperplanes of menu $\{0, a\}$

(c) Optimal hyperplanes of menu $\{0, b\}$

(d) Improvement for taking union $\{0, a, b\}$

Figure 2: Posterior-separable cost function violating IIA

4 Proof

In this section, we present a proof of our main result. To do so, we build on the connections that we illustrated in the example in Section 3.3 between our two axioms and cost functions ψ that satisfy UHP (alternatively, JDD). For a high-level discussion of the role that each axiom plays in restricting the cost of information, see Section 5.

The proof goes as follows. Section 4.1 shows that, together, IIA and IE imply that $\succsim$ has a posterior-separable representation. From this point on, all sections assume a posterior separable representation. Section 4.2 shows that the unique hyperplane property implies joint-direction differentiability and Section 4.5 shows that joint-directional differentiability implies the unique hyperplane property, proving the equivalence between (2) and (3) of Theorem 1. Then, Section 4.3 shows that IIA implies joint-directional differentiability; together with Section 4.1, this shows that (1) implies (2). Finally, Section 4.4 shows that the unique hyperplane property implies IIA and IE, showing that (3) implies (1) and finishing the proof.

4.1 Posterior-separable representation

Throughout this subsection, assume that the rationally inattentive preference $\succsim$ satisfies IIA and IE. We will show that this implies that it has a posterior separable representation.

Let $\mathcal{H}$ be the set of acts that are irrelevant to the singleton $\{0\}$, that is,

$$\mathcal{H} = \{h \in \mathcal{F} :\ 0 \sim \{0, h\}\}. \tag{6}$$

Lemma 1. *For any menu F, we have $F \subseteq \mathcal{H}$ if and only if $0 \sim F \cup \{0\}$.*

Proof. If $f \in F$, we always have $F \cup \{0\} \succsim \{0, f\} \succsim 0$, so $0 \sim F \cup \{0\}$ implies $f \in \mathcal{H}$. The other direction can be proven by induction on the size of F. Let $F = \{f_1, f_2 \ldots, f_n\} \subseteq \mathcal{H}$. If F has one element, the result follows from the definition of $\mathcal{H}$. Now let $F_n = F_{n-1} \cup \{f_n\}$ be of size n and assume that the result holds for menus of size $n-1$. Then we have $F_{n-1} \cup \{0\} \sim 0$. Since $f_n \in F_n \subseteq \mathcal{H}$, it follows that $0 \sim \{0, f_n\}$. By IIA, we must have $F_n \cup \{0\} = F_{n-1} \cup \{0, f_n\} \sim 0$, as we wanted.

□

To simplify notation, we now write

$$\langle \phi, \pi \rangle = \int \phi \, d\pi$$

for any integrable function ϕ.

Lemma 2. *The cost function c is given by*

$$c(\pi) = \sup_{\substack{F \subseteq \mathcal{H} \\ 0 \in F}} \langle \phi_F, \pi \rangle .$$

Proof. Let F be any menu. By IE, there exists an act h such that $h \sim F \sim F \cup h$. Note that, since ϕ_h is an affine function, $\phi_{F \cup h} - \phi_h$ is a piecewise linear convex function. Since u is surjective, there is a menu G, with $0 \in G$, such that $\phi_G = \phi_{F \cup h} - \phi_h$. This menu G satisfies two important properties: First,

$$V(\phi_G) = V(\phi_{F \cup h} - \phi_h) = V(\phi_{F \cup h}) - \phi_h(p_0) = V(\phi_h) - \phi_h(p_0) = 0.$$

This means that $0 \sim G \cup 0 = G$, so that $G \subseteq \mathcal{H}$. Second, for any $\pi \in \Pi(p_0)$, we have $\langle \phi_G, \pi \rangle = \langle \phi_{F \cup h} - \phi_h, \pi \rangle = \langle \phi_{F \cup h}, \pi \rangle - \phi_h(p_0)$ and therefore

$$\langle \phi_G, \pi \rangle - V(\phi_G) = \langle \phi_{F \cup h}, \pi \rangle - V(\phi_{F \cup h}) \geq \langle \phi_F, \pi \rangle - V(\phi_F),$$

since $V(\phi_F) = V(\phi_{F \cup h})$ and $\phi_{F \cup h} \geq \phi_F$. Therefore, using the formula for the cost function in Theorem 2 of De Oliveira et al. (2017),

$$c(\pi) = \sup_{F \in \mathbb{F}} \langle \phi_F, \pi \rangle - V(\phi_F) = \sup_{\substack{G \in \mathbb{F} \\ 0 \in G \\ V(\phi_G) = 0}} \langle \phi_G, \pi \rangle - V(\phi_G) = \sup_{\substack{G \subseteq \mathcal{H} \\ 0 \in G}} \langle \phi_G, \pi \rangle ,$$

since $F \subseteq \mathcal{H}$ and $0 \in F$ implies $V(\phi_F) = 0$.

□

Finally, we can prove the posterior separability of the cost function.

Lemma 3. *We can write $c(\pi) = \langle \psi, \pi \rangle$, where $\psi : \Delta(\Omega) \to \mathbb{R}$ is given by*

$$\psi(p) = \sup_{h \in \mathcal{H}} \sum_{\omega} u(h(\omega)) p(\omega) \tag{7}$$

where $\mathcal{H}$ is defined as in (6).

Proof. For any menu $F \subseteq \mathcal{H}$, we have $\phi_F \leqslant \psi$, so that

$$c(\pi) = \sup_{\substack{F \subseteq \mathcal{H} \\ 0 \in F}} \langle \phi_F, \pi \rangle \leqslant \langle \psi, \pi \rangle .$$

To show the converse inequality, fix $\epsilon > 0$ and $\pi \in \Pi(\overline{p})$ with support $p_1, \ldots, p_n$. Suppose first that $\psi(p_i) < \infty$ for $i = 1, \ldots, n$. From the definition of ψ, we can find $h_1, \ldots, h_n$ such that

$$\psi(p_i) < \langle h_i, p_i \rangle + \epsilon \text{ for } i = 1, \ldots, n.$$

Letting $F = \{0, h_1, \ldots, h_n\}$ we have

$$c(\pi) \geqslant \langle \phi_F, \pi \rangle \geqslant \sum_i \langle h_i, p_i \rangle \pi(p_i) > \langle \psi, \pi \rangle - \epsilon.$$

Since ϵ was chosen arbitrarily, this shows that $c(\pi) \geqslant \langle \psi, \pi \rangle$.

Suppose now that $\psi(p_i) = \infty$ for some i. Then there must be a sequence $(h_n)_{n=1}^{\infty}$ of acts in $\mathcal{H}$ such that $\phi_{h_n}(p_i) \to \infty$. Letting $G_n = \{0, h_n\}$ we have that

$$c(\pi) \geqslant \sup_{n \in \mathbb{N}} \langle \phi_{G_n}, \pi \rangle \geqslant \sup_{n \in \mathbb{N}} \phi_{h_n}(p_i)\pi(p_i) = \infty,$$

finishing the proof. □

4.2 Unique Hyperplane Property implies Joint-Directional Differentiability

We prove the contrapositive: if ψ does not satisfy JDD, then it does not satisfy UHP. Thus, suppose there are[8] $\{p_i\}_{i=1}^{K} \subset \operatorname{dom} \psi$, $\delta \in \mathbb{R}^M \setminus \{0\}$, and $\lambda_i \in \mathbb{R}^M$ such that

1. $\lambda_i \in \partial \bar{\psi}(\bar{p}_i)$,
2. $\lambda_i + \delta \in \partial \bar{\psi}(\bar{p}_i)$,
3. $\delta \cdot \bar{p}_i = 0$ for all i, and
4. $p_0 \in co(p_1, \ldots, p_K)$.

In order to demonstrate the violation of UHP, we will construct a menu for which it is optimal for the DM to choose information whose support is precisely $\{p_i\}_{i=1}^{K}$, and that the UHP will fail for this menu.

LEMMA 4. *Given conditions (1)-(4) above, there exists a menu H such that*

1. *there is an optimal distribution over posteriors for H with support on $\{p_i\}_{i=1}^{K}$;*
2. *$0, -\delta \in \Lambda_H$.*

Proof. For each i, let $h_i : \Omega \to X$ be such that $\sum_{\omega \in \Omega} u(h_i(\omega))p(\omega) = \lambda_i \cdot (\bar{p} - \bar{p}_i) + \bar{\psi}(\bar{p}_i)$ for every $p \in \Delta(\Omega)$ (since the image of u is $\mathbb{R}$ and the right hand side is an affine function of p, this is always possible). Defining

[8] Recall that for any $p \in \operatorname{dom}(\psi)$, we define $\bar{p} = T(p) \in Y$ (see Section 2.4).

the menu $H = \{h_1, \ldots, h_K\}$, we have

$$\phi_H(p) = \max_{h_i \in H} \sum_{\omega \in \Omega} u(h_i(\omega))p(\omega) = \max_i \lambda_i \cdot (\bar{p} - \bar{p}_i) + \bar{\psi}(\bar{p}_i).$$

Spelling out the definition of subdifferential in condition (1), we have $\lambda_i \cdot (\bar{p} - \bar{p}_i) \leq \bar{\psi}(\bar{p}) - \bar{\psi}(\bar{p}_i)$ for $i = 1, \ldots, K$. This implies that $\phi_H(p) \leq \bar{\psi}(\bar{p})$ for all $p \in \Delta(\Omega)$, which in turn implies $N_H^* \leq 0$. Moreover, $N_H^*(\bar{p}_i) = 0$ for $i = 1, \ldots, K$. Since $p_0 \in co(p_1, \ldots, p_K)$, this implies that $\text{cav}(N_H^*)(\bar{p}_0) = 0$, so $0 \in \Lambda_H$.

Similarly, spelling out the definition of subdifferential in condition (2), we have $(\lambda_i + \delta) \cdot (\bar{p} - \bar{p}_i) \leq \bar{\psi}(\bar{p}) - \bar{\psi}(\bar{p}_i)$ for $i = 1, \ldots, K$. This implies that $\phi_H(p) \leq \bar{\psi}(\bar{p}) - \delta \cdot (\bar{p} - \bar{p}_i)$ for all $p \in \Delta(\Omega)$. Since $\delta \cdot (\bar{p}_i - \bar{p}_0) = 0$ for all i, it follows that $N_H^*(\bar{p}) \leq -\delta \cdot (\bar{p} - \bar{p}_0)$ for all $\bar{p} \in Y$. Moreover, $N_H^*(\bar{p}_i) = 0 = -\delta \cdot (\bar{p}_i - \bar{p}_0)$ for $i = 1 \ldots, K$. As before, this implies that $-\delta \in \Lambda_H$. Since $\delta \neq 0$, this shows that the unique hyperplane property is not satisfied. □

4.3 IIA implies JDD

We now show that, if the preference has a posterior-separable representation and satisfies IIA, the canonical cost function ψ will satisfy JDD.

We start with a preliminary lemma regarding the subdifferentials of sets of points that are NDISD.

Lemma 5. *The set*

$$D(\{\bar{p}_i\}_{i=1}^K) := \bigcap_i \left(\partial\bar{\psi}(\bar{p}_i) - \partial\bar{\psi}(\bar{p}_i)\right) \cap \bar{p}_i^{\perp}. \tag{8}$$

is compact and convex. Moreover, $\bar{\psi}$ is non-differentiable in the direction δ at at $\{\bar{p}_i\}_{i=1}^K$ if and only if $\delta \in D(\{\bar{p}_i\}_{i=1}^K)$.

Proof. See Appendix A.3. □

Lemma 6. *Suppose $\succsim$ has a posterior separable representation with a ψ that violates JDD. Then $\succsim$ violates IIA.*

We use these properties of the set to construct a violation of IIA, generalizing the intuiting in the example in Section 3.3. That is, when JDD fails, one can find two different acts that, when added to the original menu, preserve the original respective hyperplanes, and hence the value. However, when adding *both*, the new concavification lies strictly higher. This is in contradiction to IIA, which states that it must continue to provide the original value of the menu.

Proof. By Lemma 5, the set $D(\{\bar{p}_i\}_{i=1}^K)$ is compact, convex, and contains a non-null linear functional. Let δ be an extreme point in this set. Note that if $\delta \in D(\{\bar{p}_i\}_{i=1}^K)$, then so is $-\delta$.

Let H be a set of acts as given by Lemma 4, so that $0, -\delta \in \Lambda_H$. Let $\epsilon \in \mathbb{R}^M$ be a sufficiently small vector such that

1. $\bar{p}_0 + \epsilon, \bar{p}_0 - \epsilon \in Y$;

2. $\delta \cdot \epsilon > 0$.

Such an ϵ exists because the orthogonal complement of $D(\{p_i\}_{i=1}^K)$ does not have full rank. Let $\alpha \in \partial\psi(\bar{p}_0+\epsilon)$ and $\beta+\delta \in \partial\bar{\psi}(\bar{p}_0-\epsilon)$. Then

$$\alpha \cdot \bar{p} - \bar{\psi}(\bar{p}) \leqslant \alpha \cdot (\bar{p}_0 + \epsilon) - \bar{\psi}(\bar{p}_0 + \epsilon) = 0$$
$$(\beta+\delta) \cdot \bar{p} - \bar{\psi}(\bar{p}) \leqslant (\beta+\delta) \cdot (\bar{p}_0 - \epsilon) - \bar{\psi}(\bar{p}_0 - \epsilon) = 0$$

We now define two other menus:

$$F = \{f_\alpha\} \cup H$$
$$G = \{f_\beta\} \cup H.$$

where f_α is defined so that

$$\sum_{\omega\in\Omega} u(f_\alpha(\omega))p(\omega) = \alpha \cdot (\bar{p} - \bar{p}_0 - \epsilon) + \bar{\psi}(\bar{p}_0 + \epsilon)$$

holds for every $p \in \mathrm{dom}(\psi)$, and similarly for f_β,

$$\sum_{\omega\in\Omega} u(f_\beta(\omega))p(\omega) = \beta \cdot (\bar{p} - \bar{p}_0 + \epsilon) + \bar{\psi}(\bar{p}_0 - \epsilon) + \delta \cdot \epsilon.$$

Notice that $0 \in \Lambda_H$ and since $\alpha\cdot(\bar{p}-\bar{p}_0-\epsilon)+\bar{\psi}(\bar{p}_0+\epsilon)-\bar{\psi}(\bar{p}) \leqslant 0$, it remains in Λ_F, and so the concavification still yields $V(\phi_F) = 0$ by (5). Likewise, $-\delta \in \Lambda_H$ and since $\beta\cdot(\bar{p}-\bar{p}_0+\epsilon)+\bar{\psi}(\bar{p}_0-\epsilon)-\bar{\psi}(\bar{p})+\delta\cdot\epsilon \leqslant -\delta\cdot(\bar{p}-\bar{p}_0)$, it remains in Λ_G, and so the concavification still yields $V(\phi_G) = 0$. Thus, all three menus — F,G, and H — give the same value of zero.

Now consider $\gamma \in \mathbb{R}^M$ and $k \in \mathbb{R}$ such that $\gamma \cdot \bar{p} + k \geqslant N^*_{F\cup G}(\bar{p})$ for all $\bar{p}$. We must have

$$\gamma \cdot (\bar{p}_0 + \epsilon) + k \geqslant \alpha \cdot (\bar{p}_0 + \epsilon - \bar{p}_0 - \epsilon) + \bar{\psi}(\bar{p}_0 + \epsilon) - \bar{\psi}(\bar{p}_0 + \epsilon) = 0$$
$$\gamma \cdot (\bar{p}_0 - \epsilon) + k \geqslant \beta \cdot (\bar{p}_0 - \epsilon - \bar{p}_0 + \epsilon) + \bar{\psi}(\bar{p}_0 - \epsilon) + \delta \cdot \epsilon - \bar{\psi}(\bar{p}_0 - \epsilon) = \delta \cdot \epsilon > 0.$$

Together, these inequalities imply that $\gamma \cdot \bar{p}_0 \geq \frac{1}{2}\delta \cdot \epsilon > 0$, so whatever is the optimal hyperplane for $F \cup G$, it must get a value strictly greater than zero at $\bar{p}_0$, meaning that $V(\phi_{F\cup G}) > 0$. This shows that our preference violates IIA. ☐

4.4 Unique Hyperplane Property implies IIA and IE

We now show that if a preference has a posterior-separable representation and satisfies the Unique Hyperplane Property, then it must satisfy IIA and IE. Let F and G be menus such that $F\cap G \sim F \sim G$. Assuming that the unique hyperplane property is satisfied, we may write $\Lambda_F = \{\lambda_F\}$, $\Lambda_G = \{\lambda_G\}$, and $\Lambda_{F\cap G} = \{\lambda_{F\cap G}\}$. Since $F \cap G \sim F \sim G$, and the optimal posteriors for $F \cap G$ are also feasible for F and G, the optimal hyperplane for $F \cap G$ must also be an optimal hyperplane for both F and G. By the unique hyperplane property, the optimal hyperplanes must therefore identical: $\lambda_F = \lambda_{F\cap G} = \lambda_G$. This implies that for all $f \in F \cup G$ and $p \in \Delta(\Omega)$, $\lambda_F \cdot (\bar{p} - \bar{p}_0) + V(\phi_F) \geqslant \sum_\omega u(f(\omega))p(\omega) - \psi(p)$ so that $\lambda_F \cdot (\bar{p} - \bar{p}_0) + V(\phi_F) \geqslant N^*_{F\cup G}(\bar{p})$. and so λ_F is the (unique) optimal hyperplane for $F \cup G$.

To show that IE is satisfied, let λ_F be the optimal hyperplane for menu F. Defining act h such that $\mathbf{M}_T^{\intercal} h = \lambda_F + \partial\bar{\psi}(\bar{p}_0)$, where $\mathbf{M}_T$ is the matrix representing linear transformation T, and $\partial\bar{\psi}(\bar{p}_0)$ is a singleton by joint directional differentiability, we get that $\mathbf{M}_T^\intercal h \cdot \bar{p} - \bar{\psi}(\bar{p}) \leq \lambda_F \cdot \bar{p}, \forall \bar{p} \in Y$. □

4.5 Joint-directional Differentiability implies the Unique Hyperplane Property

Suppose the preference has a posterior-separable representation where ψ is joint-directional differentiable. Let $F := \{f_i\}_{i=1}^K$ be an arbitrary menu. For each i, let $a_i \in \mathbb{R}^M$ and $b_i \in \mathbb{R}$ be such that $\sum_{\omega\in\Omega} u(f_i(\omega))p(\omega) = a_i \cdot \bar{p} + b_i$ for all $p \in \operatorname{dom}\psi$. Let $\{p_i\}_{i=1}^I$ be the support of an optimal distribution over posteriors for F, so $p_0 \in co\{p_i\}_{i=1}^I$. We can assume, without loss of generality, that no act in F is optimal for more than one posterior belief; for simplicity, we label the acts so that f_i is optimal for p_i for $I \leq K$ (relabeling any unchosen acts to have $i > K$). Then there exists $\lambda \in \Lambda_F$ such that, for $i = 1, \ldots, I$,

$$N_F(\bar{p}_i) = a_i \cdot \bar{p}_i + b_i - \bar{\psi}(\bar{p}_i) = \lambda \cdot (\bar{p}_i - \bar{p}_0) + V(\phi_F)$$

and

$$a_i \cdot \bar{p} + b_i - \bar{\psi}(\bar{p}) \leqslant \lambda \cdot (\bar{p} - \bar{p}_0) + V(\phi_F)$$

for all $\bar{p} \in Y$. Subtracting the equality from the inequality, we get

$$(a_i + \lambda) \cdot (\bar{p} - \bar{p}_i) \leqslant \bar{\psi}(\bar{p}) - \bar{\psi}(\bar{p}_i),$$

which means that $\lambda_i := a_i + \lambda \in \partial\bar{\psi}(\bar{p}_i)$. If there existed a second hyperplane $\hat{\lambda}$, we could repeat the same argument, and letting $\delta = \lambda - \hat{\lambda}$, we would get that

$$\lambda_i + \delta = \hat{\lambda}_i \in \partial\bar{\psi}(\bar{p}_i)$$

$$\delta \cdot (\bar{p}_i - \bar{p}_0) = \lambda_i \cdot (\bar{p}_i - \bar{p}_0) - \hat{\lambda}_i \cdot (\bar{p}_i - \bar{p}_0)$$
$$= [N_F^*(\bar{p}_i) - V(\phi_F)] - [N_F^*(\bar{p}_i) - V(\phi_F)] = 0$$

So, $\bar{\psi}$ would be NDISD at $\{\bar{p}_i\}_{i=1}^I$, contradicting our assumption. □

5 Discussion of Main Theorem

As was noted when we introduced canonical costs, it is without loss to write $c(\pi) = \sup_{\psi\in\Psi} \int \psi d\pi$. For this discussion, we explore the implications of our axioms when the set Ψ is finite.[9] Then, we not only have $\psi(p_0) \leq 0$ for all $\psi \in \Psi$, but also have equality for at least one $\psi \in \Psi$. We may also assume that Ψ is minimal.

For any finite menu H, let $\pi_H \in \Pi(p_0)$ be an optimal information choice given that menu. Define $\psi_H \in \arg\max_{\psi\in\Psi} \int \psi d\pi_H$; by the minimality of Ψ, there will be at least one menu H for which ψ_H is the unique maximizer.

IE, Ignorance Equivalence, comes to rule out that there exist $\psi^* \in \Psi$ such that $\psi^*(p_0) < 0$. To see this, suppose that such a $\psi^* \in \Psi$ exists. Let $\Psi_0 = \{\psi \in \Psi | \psi(p_0) = 0\}$, which is nonempty by Proposition 2. By the minimality of Ψ, there is some menu F^* such that $c(\pi_{F^*}) = \int \psi^* d\pi_{F^*}$ and, for all $\psi \neq \psi^*$, $c(\pi_{F^*}) < \int \psi d\pi_{F^*}$.

[9]Notice that, by definition of supremum, one can approximate the cost function c by a finite set $\hat{\Psi} \subset \Psi$, defining a cost function $\hat{c}(\pi) \equiv \max_{\psi\in\hat{\Psi}} \int \psi d\pi$, as follows. In a case where $c(\pi)$ is finite, then there exists a sufficiently large, finite $\hat{\Psi} \subset \Psi$ such that $c(\pi) < \hat{c}(\pi) + \epsilon$. if $c(\pi) = \infty$, then for sufficiently large $\hat{\Psi}$, π is dominated by δ_{p_0}.

Let $\delta := \min\{-\psi^*(p_0), \min_{\psi\neq\psi^*} c(\pi_{F^*}) - \int \psi \, d\pi_{F^*}\}$.

For any h such that $h \sim F^*$, one will have $\pi_h = \delta_{p_0}$ and hence $\psi_h \in \Psi_0$. Consider now the menu $F^* \cup \{h\}$. By choosing $\pi' := \frac{1}{2}\pi_{F^*} + \frac{1}{2}\delta_{p_0}$, one can save on information costs: from the fact that $\psi^* \notin \Psi_0$ and the representation of c convex as in Proposition 2, there exists $\psi' \in \Psi$ such that

$$c(\pi') = \int \psi' d\pi'$$

$$\leq \frac{1}{2}\int \psi^* d\pi_{F^*} + \frac{1}{2}\psi_h(p_0) - \frac{1}{2}\delta$$

$$\leq \frac{1}{2}c(\pi_{F^*}) - \frac{1}{2}\delta \tag{9}$$

However, by the monotonicity of the indirect utility from the decisions in the menu size, due to the preference for flexibility,

$$\int \phi_{F^*\cup\{h\}} d\pi_{F^*\cup\{h\}} \geq \frac{1}{2}\int \phi_{F^*} d\pi_{F^*} + \frac{1}{2}\phi_h(p_0), \tag{10}$$

Therefore,

$$V(\phi_{F^*\cup\{h\}}) \geq \frac{1}{2}\int \phi_{F^*} d\pi_{F^*} + \frac{1}{2}\phi_h(p_0) - c(\pi')$$

$$\geq \frac{1}{2}[\int \phi_{F^*} d\pi_{F^*} - \int \psi^* d\pi_{F^*}] - \frac{1}{2}[\phi_h(p_0)] + \frac{1}{2}\delta$$

$$= V(\phi_{F^*}) + \frac{1}{2}\delta$$

where the last equality is from $h \sim F^*$. So, $F^* \cup \{h\} \succ F^*$, violating IE.

On the other hand, there is no problem with IE if $\Psi = \Psi_0$. In this case, for a given ψ_{F^*}, the act h is consistent with ψ_{F^*} itself, and so there are no savings of information costs to be had by randomizing.

The presence of IE allows for IIA to have bite. Without IE, it is unclear whether there will actually be any menus F, G with $\psi_F \neq \psi_G$ such that $F \cap G \neq \emptyset$; if that were so, then in such cases, IIA holds vacuously for F, G. However, with IE, we know that $\psi_F(p_0) = \psi_G(p_0) = 0$. By IE, there exist h, h' such that $h \sim F \sim F \cup \{h\}$ and $h' \sim G \sim G \cup \{h'\}$. Since adding an affine function α to the payoffs of all acts in the menu does not change the optimal choice of information, and changes the value of the menu by $\alpha \cdot p_0$, one can replace $G \cup \{h'\}$ with $\hat{G} \cup \{h\}$ by letting $\alpha = h - h'$, such that $\psi_G = \psi_{\hat{G}}$ and $F \sim G$. As a result, under IE, if there are $\psi_F \neq \psi_G$, then there are corresponding F, G such $h \sim F \sim F \cup \{h\} \sim G \cup \{h\} \sim G$. Note that $\psi_F = \psi_{F\cup\{h\}}$ and $\psi_G = \psi_{G\cup\{h\}}$ since π_F and π_G are optimal information choices for $F \cup \{h\}$ and $G \cup \{h\}$, respectively.

Given such F, G, then by taking $F \cup G \cup \{h\}$, IIA dictates that $F \sim F \cup G \cup \{h\}$. However, if $\psi_F \neq \psi_G$, this is not the case: one can save on the information cost by randomizing $\frac{1}{2}\pi_F + \frac{1}{2}\pi_G$, while keeping the expected utility from the decisions the same, similarly to the cost saving/expected utility preservation due to IE in equations (9) and (10). So, $\psi_F = \psi_G$.

6 Uniform Posterior-Separable Costs

We now allow for the agent's prior to vary. For each possible prior $p_0 \in \Delta(\Omega)$, we denote the cost of information by $c_{p_0} : \Pi(p_0) \to \mathbb{R} \cup \{\infty\}$. For each $\pi \in \Delta(\Delta(\Omega))$, we can deduce the corresponding prior p_0 by taking the expectation of π, so that $\pi \in \Pi(p_0)$. Thus, we may write the cost function for all priors under a single notation $c : \Pi \to \mathbb{R}$. We call c canonical if, for each p_0, c_{p_0} is canonical. We now consider the following condition that allows for the costs to be "the same" across menus.

Definition 7. *The canonical cost function* $c : \Delta(\Delta(\Omega)) \to \mathbb{R} \cup \{\infty\}$ *is* prior invariant *with respect to the set* Ψ *of convex functions that* represents c *if*

$$c(\pi) = \sup_{\psi \in \Psi} \int \psi \, d\pi - \sup_{\psi' \in \Psi} \psi' \left(\int p \, d\pi \right).$$

When Ψ *is a singleton, we say that* c *is* uniformly posterior separable*.*

Note that if c is canonical then, by Proposition 2, for every p_0 there exists a set of convex functions Ψ_{p_0} such that

$$c(\pi) = \sup_{\psi \in \Psi_{p_0}} \int \psi \, d\pi = \sup_{\psi \in \Psi_{p_0}} \int \psi \, d\pi - \max_{\psi' \in \Psi_{p_0}} \psi'(p_0).$$

Prior-invariance imposes that the same set Ψ applies to every p_0. Note that we need to subtract the second term to guarantee that c_{p_0} is grounded for every p_0.

The quintessential uniformly posterior-separable cost is the mutual information (Sims, 2003; Caplin, Dean, and Leahy, 2022). Denoting the entropy of a distribution by $H(p) = -\sum_\omega p(\omega) \ln(p(\omega))$, mutual information is given by the expected reduction in entropy,

$$c_{p_0}(\pi) = H(p_0) - \int H(p) \, \pi(dp),$$

which matches the formula in definition 7 by setting $\Psi = \{-H\}$.

Theorem 2. *For a fixed prior* p_0*, let* Ψ *be a set of convex functions representing* c_{p_0} *satisfying the conditions of Proposition 2. Let the canonical cost function* c *be prior-invariant with respect to* Ψ*. If, for each prior, the preference over menus corresponding to* c *satisfies IIA and IE, then* c *is uniformly posterior separable and represented by some* $\psi : \Delta(\Omega) \to \mathbb{R} \cup \{\infty\}$ *that is differentiable at all* $p \in \mathrm{ri}(\mathrm{dom}(\psi))$*.*

Proof. By Theorem 1, the cost function c_{p_0} is posterior separable. Since Ψ is minimal for c_{p_0}, we must have $\Psi = \{\psi\}$ for some convex $\psi : \Delta(\Omega) \to \mathbb{R} \cup \{\infty\}$. Since c is prior-invariant, we must have that, for all priors $q \in \Delta(\Omega)$ and $\pi \in \Pi(q)$,

$$c_q(\pi) = \sup_{\psi' \in \Psi} \int \psi'(p) d\pi - \sup_{\psi'' \in \Psi} \psi''(q) = \int \psi(p) d\pi - \psi(q),$$

so c is uniformly posterior separable.

To see that ψ must be differentiable, suppose, to the contrary, that it is not differentiable at some $q \in \mathrm{ri}(\mathrm{dom}(\psi))$. Fixing the prior to be q, we have, by Theorem 1, that the preference $\succsim$, satisfying IIA and IE, has a posterior separable representation with a canonical measure of uncertainty $\hat{\psi}$ that satisfies joint-directional differentiability. Thus for every $\pi \in \Pi(q)$,

$$\int \hat{\psi} \, d\pi = \int \psi \, d\pi - \psi(q) = \int [\psi(p) - \psi(q)] \, \pi(dp).$$

By Lemma 7 in Appendix A.2, there exists ξ such that

$$\hat{\psi}(p) = \psi(p) - \psi(q) + \xi \cdot (p - q).$$

Since $\hat{\psi}$ satisfies joint-directional differentiability, it must be differentiable at q (simply pick $K = 1$ and $\{p_i\}_{i=1}^{K} = \{q\}$ and we trivially have $q \in co(\{q\})$). This implies that ψ is differentiable at q, contradicting its non-differentiability. $\square$

References

Aumann, Robert J and Michael Maschler. 1995. *Repeated games with incomplete information.* MIT press.

Blackwell, David. 1951. "Comparison of experiments." In *Proceedings of the second Berkeley symposium on mathematical statistics and probability*, vol. 1. 26.

Caplin, Andrew and Mark Dean. 2015. "Revealed preference, rational inattention, and costly information acquisition." *American Economic Review* 105 (7):2183–2203.

Caplin, Andrew, Mark Dean, and John Leahy. 2022. "Rationally inattentive behavior: Characterizing and generalizing Shannon entropy." *Journal of Political Economy* 130 (6):1676–1715.

de Oliveira, Henrique. 2014. "Axiomatic foundations for entropic costs of attention." Tech. rep., Technical report, Mimeo.

De Oliveira, Henrique, Tommaso Denti, Maximilian Mihm, and Kemal Ozbek. 2017. "Rationally inattentive preferences and hidden information costs." *Theoretical Economics* 12 (2):621–654.

Denti, Tommaso. 2022. "Posterior separable cost of information." *American Economic Review* 112 (10):3215–59.

Denti, Tommaso, Massimo Marinacci, and Aldo Rustichini. 2021. "Experimental Cost of Information." *Working paper* .

Ergin, Haluk and Todd Sarver. 2010. "A unique costly contemplation representation." *Econometrica* 78 (4):1285–1339.

Gleyze, Simon and Agathe Pernoud. 2023. "Informationally simple incentives." *Journal of Political Economy* 131 (3):802–837.

Hébert, Benjamin and Jennifer La'O. 2023. "Information acquisition, efficiency, and nonfundamental volatility." *Journal of Political Economy* 131 (10):2666–2723.

Hébert, Benjamin and Michael Woodford. 2021. "Neighborhood-based information costs." *American Economic Review* 111 (10):3225–3255.

Kamenica, Emir and Matthew Gentzkow. 2011. "Bayesian Persuasion." *American Economic Review* 101 (October):2590–2615.

Lipnowski, Elliot, Laurent Mathevet, and Dong Wei. 2020. "Attention management." *American Economic Review: Insights* 2 (1):17–32.

Lipnowski, Elliot and Doron Ravid. 2022. "Predicting choice from information costs." *arXiv preprint arXiv:2205.10434* .

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

Mensch, Jeffrey. 2022. "Screening inattentive buyers." *American Economic Review* 112 (6):1949–1984.

Mensch, Jeffrey and Komal Malik. 2024. "Posterior-Mean Separable Costs of Information Acquisition." *arXiv preprint arXiv:2311.09496* .

Müller-Itten, Michele, Roc Armenter, and Zachary Stangebye. 2023. "Rational inattention via ignorance equivalence." .

Pomatto, Luciano, Philipp Strack, and Omer Tamuz. 2023. "The cost of information: The case of constant marginal costs." *American Economic Review* 113 (5):1360–1393.

Rockafellar, Ralph Tyrell. 1970. *Convex Analysis*. Princeton: Princeton University Press. URL https://doi.org/10.1515/9781400873173.

Sims, Christopher A. 2003. "Implications of rational inattention." *Journal of monetary Economics* 50 (3):665–690.

Yang, Ming. 2020. "Optimality of debt under flexible information acquisition." *The Review of Economic Studies* 87 (1):487–536.

Yoder, Nathan. 2022. "Designing incentives for heterogeneous researchers." *Journal of Political Economy* 130 (8):2018–2054.

A Appendix

A.1 Proof of Proposition 2

By De Oliveira et al. (2017), Theorem 2, the cost of information can be written as

$$c(\pi) = \sup_{F \in \mathbb{F}} \int \phi_F(p) d\pi(p) - V(F)$$

Notice that for each F, $\phi_F - V(F)$ is a convex function, and so one can let $\Psi = \{\phi_F - V(F) : F \in \mathbb{F}\}$ to establish

$$c(\pi) = \sup_{\psi \in \Psi} \int \psi d\pi$$

Write $\Psi' \sim \Psi$ to denote the equivalence relation that says that Ψ' and Ψ represent the same cost function c. We now show that there exists a $\Psi' \sim \Psi$ satisfying properties 1 and 2 in the proposition. To do so, we divide in two cases:

1. Suppose that there exists a finite $\Psi' \sim \Psi$. If Ψ' is minimal, we are done; if not, there exists a Ψ'' with $|\Psi''| < |\Psi|$ such that $\Psi'' \sim \Psi' \sim \Psi$. We can then repeat the same argument, until we reach a minimal Ψ^*, which must happen in a finite number of steps since Ψ' was finite. Since Ψ^* is finite, the maximum must always be achieved, and in particular $c(\delta_{p_0}) = \max_{\psi \in \Psi^*} \psi(p_0) = 0$.

2. Suppose that there is no finite $\Psi' \sim \Psi$. Then Ψ is already minimal. To satisfy condition (1), note that, since c is grounded,

$$0 = c(\delta_{p_0}) = \sup_{\psi \in \Psi} \psi(p_0),$$

and since c is also Blackwell monotone, $c(\pi) \geqslant 0$ for all $\pi \in \Pi(p_0)$. This means that if we add $\psi_0 \equiv 0$ to Ψ, we get

$$\sup_{\psi \in \Psi \cup \{\psi_0\}} \int \psi \, d\pi = \max \left\{ \sup_{\psi \in \Psi} \int \psi \, d\pi, 0 \right\} = \max\{c(\pi), 0\} = c(\pi)$$

so $\Psi \cup \{\psi_0\} \sim \Psi$, and

$$c(\delta_{p_0}) = \psi_0(p_0) = 0 = \max_{\psi \in \Psi \cup \{\psi_0\}} \psi(p_0),$$

which is condition (1). Since $|\Psi \cup \{\psi_0\}| = |\Psi| = \infty$, condition (2) is also satisfied by $\Psi \cup \{\psi_0\}$.

A.2 Proof of Proposition 3

Let c be canonical and represented by ψ.
(1) It follows from Blackwell monotonicity that ψ must be convex (see Lipnowski and Ravid (2022), Lemma 24).
(2) From groundedness, it is immediate that $\psi(p_0) = 0$.
(4) We now construct a $\hat{\psi}$ that represents the same cost of information as ψ, satisfying $p_0 \in \operatorname{ri} \operatorname{dom} \hat{\psi}$. If $p_0 \in \operatorname{ri} \operatorname{dom} \psi$, we simply let $\hat{\psi} = \psi$. So suppose $p_0 \notin \operatorname{ri} \operatorname{dom} \psi$ and let $d = \dim(\operatorname{dom} \psi)$ denote the dimension of $\operatorname{dom} \psi$. Since p_0 is on the boundary of $\operatorname{dom} \psi$, by Rockafellar (1970), Theorem 11.6, there exists a supporting hyperplane, defined by some vector λ, such that, letting

$$X := \{p \in \Delta(\Omega) : \lambda \cdot p \geq \lambda \cdot p_0\}$$

we have that for all $p \in X$,

$$\lambda \cdot p > \lambda \cdot p_0 \implies \psi(x) = \infty,$$

while $X \cap \mathrm{ri}(\mathrm{dom}(\psi)) = \emptyset$, i.e. X only intersects with $\mathrm{dom}(\psi)$ on the boundary of the latter.

Now let

$$\hat{\psi}(p) = \begin{cases} \psi(p), & \text{if } p \in X \\ \infty, & \text{otherwise} \end{cases}$$

That is, $\hat{\psi}$ replaces $\psi(p)$ with ∞ for all $p \notin X$. In particular, $\hat{\psi}(p) = \infty$, $\forall p \in \mathrm{ri}(\mathrm{dom}(\psi))$. We shall show that $c(\pi) = \int \psi \, d\pi = \int \hat{\psi} \, d\pi$ for every $\pi \in \Pi(p_0)$. Indeed, if $\mathrm{supp}(\pi) \subset X$, this is obviously true. When $\mathrm{supp}(\pi) \not\subset X$, by Bayes' rule then, with positive probability according to π, $\lambda \cdot p > \lambda \cdot p_0$, and therefore $\psi(p) = \infty$. Consequently, $c(\pi) = \infty = \int \hat{\psi} \, d\pi$.

If $p_0 \in \mathrm{ri}(\mathrm{dom}(\hat{\psi}))$, we found our desired $\hat{\psi}$. Otherwise, note that $\dim(\mathrm{dom}(\hat{\psi})) < \dim(\mathrm{dom}(\psi))$ and repeat the procedure until $p_0 \in \mathrm{ri}(\mathrm{dom}(\hat{\psi}))$ or the dimension is zero (in which case the domain is just p_0 and the cost of information was trivial).

(3) Let $T : \mathrm{aff}(\mathrm{dom}\, \hat{\psi}) \to \mathbb{R}^M$ be a bijective affine transformation as in Section 2.4. Since $\hat{\psi}$ is convex, so is $T^*\hat{\psi}$, so there exists a $\lambda \in \partial T^*\hat{\psi}$, that is, $\lambda \cdot (\bar{p} - \bar{p}_0) \leqslant T^*\hat{\psi}(\bar{p}) = \hat{\psi}(p)$ for all $p \in \mathrm{dom}\, \hat{\psi}$. Let $\tilde{\psi}(p) = \hat{\psi}(p) - \lambda \cdot (\bar{p} - \bar{p}_0)$. Then $\tilde{\psi} \geqslant 0$ and it inherits properties (1), (2), and (4) from $\hat{\psi}$, so $\tilde{\psi}$ is canonical. Also note that, for all $\pi \in \Pi(p_0)$,

$$\int \tilde{\psi} \, d\pi = \int \left[\hat{\psi}(p) - \lambda \cdot (T(p) - T(p_0)) \right] \, \pi(dp) = \int \hat{\psi} \, d\pi - \lambda \cdot (T(p_0) - T(p_0)) = c(\pi).$$

Before proving the uniqueness of the canonical representation, we prove the following lemma.

Lemma 7. *If ψ and ψ' are canonical and $\int \psi \, d\pi = \int \psi' \, d\pi$ for all $\pi \in \Pi(p_0)$ then there exists a $\xi \in \mathbb{R}^{|\Omega|}$ such that $\psi(p) = \psi'(p) + \xi \cdot (p - p_0)$ for all $p \in \Delta(\Omega)$.*

Proof. We first prove that $\mathrm{dom}\, \psi = \mathrm{dom}\, \psi'$. Let $p \in \mathrm{dom}\, \psi$ be arbitrary. Then $p_0 + t(p - p_0) \in \mathrm{aff}\,\mathrm{dom}\, \psi$ for all $t \in \mathbb{R}$. Since $p_0 \in \mathrm{ri}\,\mathrm{dom}\, \psi$, we can find an $\epsilon > 0$ small enough that $p' := p_0 - \epsilon(p - p_0) \in \mathrm{dom}\, \psi$. Then p_0 is in the convex hull of $\{p, p'\}$, which means we can find a $\pi \in \Pi(p_0)$ with support $\{p, p'\} \subset \mathrm{dom}\, \psi$. Then $c(\pi) = \int \psi \, d\pi < \infty$, which means that $\int \psi' \, d\pi < \infty$ as well, which can only happen if $p \in \mathrm{dom}\, \psi'$. Hence $\mathrm{dom}\, \psi \subset \mathrm{dom}\, \psi'$ and, by symmetry, $\mathrm{dom}\, \psi = \mathrm{dom}\, \psi'$. This proves that $\psi(p) = \psi'(p) + \xi \cdot (p - p_0)$ for all $p \notin \mathrm{dom}\, \psi$ regardless of ξ since the finite term $\xi \cdot (p - p_0)$ becomes irrelevant.

Now, for $p \in \mathrm{dom}\, \psi$, let $\zeta(p) = \psi(p) - \psi'(p)$. Since ψ and ψ' are canonical, $\zeta(p_0) = 0$. By Theorem 1.5 in Rockafellar (1970), the proof will be finished if we show that ζ is an affine function. To that end, let $\alpha \in [0, 1]$ and $x, y \in \mathrm{dom}\, \psi$. Suppose first that $\alpha x + (1 - \alpha)y = p_0$. Then, letting $\pi \in \Pi(p_0)$ have support $\{x, y\}$, we get

$$0 = \int \zeta \, d\pi = \alpha\zeta(x) + (1 - \alpha)\zeta(y).$$

Since $\zeta(\alpha x + (1 - \alpha)y) = \zeta(p_0) = 0$, we have $\zeta(\alpha x + (1 - \alpha)y) = \alpha\zeta(x) + (1 - \alpha)\zeta(y)$.

Now suppose that $\alpha x + (1 - \alpha)y \neq p_0$. Let $\epsilon > 0$ be small enough that $z = p_0 - \epsilon(\alpha x + (1 - \alpha)y - p_0) \in \mathrm{dom}\, \psi$ (recall that $p_0 \in \mathrm{ri}\,\mathrm{dom}\, \psi$). Letting $t = (1 + \epsilon)^{-1} \in (0, 1)$, we have that $p_0 = tz + (1 - t)(\alpha x + (1 - \alpha)y)$. Thus there is a $\pi \in \Pi(p_0)$ with support on $\{z, \alpha x + (1 - \alpha)y\}$ so that

$$0 = \int \zeta \, d\pi = t\zeta(z) + (1 - t)\zeta(\alpha x + (1 - \alpha)y). \tag{11}$$

Similarly, we may write $p_0 = tz + (1-t)\alpha x + (1-t)(1-\alpha)y$ so there is a $\pi' \in \Pi(p_0)$ with support $\{x, y, z\}$ putting probability t on z. Hence

$$0 = \int \zeta \, d\pi' = t\zeta(z) + (1-t)\alpha\zeta(x) + (1-t)(1-\alpha)\zeta(y). \tag{12}$$

Putting together equations eq. (11) and eq. (12), we get that $\zeta(\alpha x + (1-\alpha)y) = \alpha\zeta(x) + (1-\alpha)\zeta(y)$. Thus, we have shown that ζ is affine, which finishes the proof. □

Finally, if ψ is differentiable in the directions of its domain at p_0, so are $\hat{\psi}$ and $\tilde{\psi}$. If ψ' were another canonical measure of uncertainty representing the same cost of information, then by Lemma 7, there would be a ξ such that $\tilde{\psi}(p) = \psi'(p) + \xi \cdot (p - p_0)$. Also, by the proof of Lemma 7, $\tilde{\psi}$ and ψ' must have the same domain, so we may use the same bijective affine transformation T for both. Since $T^*\psi'(p) \geqslant 0$ and $T^*\tilde{\psi}(p_0) = 0$, this would imply that $T^*\tilde{\psi}(p) - T^*\tilde{\psi}(p_0) \geqslant \bar{\xi} \cdot (\bar{p} - \bar{p}_0)$, or $\bar{\xi} \in \partial T^*\tilde{\psi}$. But since $\tilde{\psi}$ is differentiable in the directions of its domain at p_0, $\bar{\xi} = 0$ must be the only element in the subdifferential, meaning that $T^*\tilde{\psi} = T^*\psi'$, which implies that $\tilde{\psi} = \psi'$.

A.3 Proof of Lemma 5

Notice that $0 \neq \delta \in D(\{\bar{p}_i\}_{i=1}^K)$ if and only if $\delta \in \partial\bar{\psi}(\bar{p}_i) - \partial\bar{\psi}(\bar{p}_i)$ and $\delta \cdot (\bar{p}_i - \bar{p}_0) = 0$ for each p_i, $i = 1, \ldots, K$, which is precisely the definition of ψ being non-differentiable in the direction δ. By Rockafellar (1970, Theorem 23.2), $\partial\bar{\psi}(\bar{p}_i)$ is closed and convex; since this is preserved under subtraction and intersection, so is $D(\{\bar{p}_i\}_{i=1}^K)$. To show that $D(\{\bar{p}_i\}_{i=1}^K)$ is compact, it remains to show that $\partial\bar{\psi}(\bar{p}_i)$ is bounded. By condition (2) of non-differentiability in the same direction, we have $(\lambda + \delta)(\bar{p}_i - \bar{p}_0) = 0$. Using Lemma 4, fix a menu H for which the posteriors $\{p_i\}_{i=1}^K$ are optimal. Notice that in this case, if $\lambda, \lambda' \in \Lambda_H$, then since for $\hat{\lambda} \in \{\lambda, \lambda'\}$

$$\lambda \cdot (\bar{p}_i - \bar{p}_0) + V(\phi_H) = N_H^*(\bar{p}_i), \forall i$$

it follows that $(\lambda - \lambda') \cdot (\bar{p}_i - \bar{p}_0) = 0$. Therefore,

$$D(\{\bar{p}_i\}_{i=1}^K) = \Lambda_H - \Lambda_H$$

and so $D(\{\bar{p}_i\}_{i=1}^K)$ is compact if Λ_H is. Then, since $\delta \cdot (\bar{p}_i - \bar{p}_0) = 0$, $V(\phi_H) = \lambda \cdot (\bar{p}_0 - \bar{p}_i) + N_F^*(\bar{p}_i) = [\lambda + \delta] \cdot (\bar{p}_0 - \bar{p}_i) + N_F^*(\bar{p}_i)$, for all $i \in \{1, ..., K\}$, and therefore $\lambda + \delta$ is also an optimal hyperplane. By way of contradiction, suppose Λ_H were not compact, then by Rockafellar (1970, Theorem 8.4), there would exist $\lambda, \delta \neq 0$ such that, for all $t > 0$, $\lambda + t\delta \in \Lambda_H$. Furthermore, as $D(\{\bar{p}_i\}_{i=1}^K)$ has dimension at least 1, its orthogonal complement does not have full rank (Rockafellar, 1970, p. 5). As a result, for all $\epsilon > 0$, there would exist p such that $\|\bar{p} - \bar{p}_0\| < \epsilon$, $\delta \cdot (\bar{p} - \bar{p}_0) < 0$, and $\phi(p) - \psi(p) \leq V(\phi_H) + [\lambda + t\delta] \cdot (\bar{p} - \bar{p}_0)$, $\forall t > 0$, which can only happen if $\psi(p) = \infty$. Thus, there would be a sequence $\{\hat{p}_j\}_{j=1}^\infty \subset \mathrm{dom}(\psi)$ such that $\lim_{j\to\infty} \hat{p}_j = p_0$ and $\psi(\hat{p}_j) = \infty, \forall j$, contradicting the assumption that $p_0 \in \mathrm{ri}(\mathrm{dom}(\psi))$ for ψ canonical with $\dim(\mathrm{dom}(\psi)) = M$.