Predicting Choice from Information Costs *

Elliot Lipnowski[†]
Columbia University

Doron Ravid[‡]
University of Chicago

May 20, 2022

Abstract

An agent acquires a costly flexible signal before making a decision. We explore the degree to which knowledge of the agent's information costs helps predict her behavior. We establish an impossibility result: learning costs alone generate no testable restrictions on choice without also imposing constraints on actions' state-dependent utilities. By contrast, for most utility functions, knowing both the utility and information costs enables a unique behavioral prediction. Finally, we show that for smooth costs, most choices from a menu uniquely pin down the agent's decisions in all submenus.

*We would like to thank Ben Brooks, Tommaso Denti, Mira Frick, Daniel Gottlieb, Ryota Iijima, Navin Kartik, R. Vijay Krishna, Ce Liu, Daniel Martin, Stephen Morris, Pietro Ortoleva, and Kareen Rozen for their comments. Tianhao Liu provided excellent research assistance.

[†] e.lipnowski@columbia.edu

[‡] dravid@uchicago.edu

1. Introduction

Sparked by Sims' (1998; 2003; 2006) studies on rational inattention, the last two decades have seen a growing interest in models of costly flexible learning. Compared with the traditional framework of decision-making under uncertainty, these models postulate that the agent's behavior depends on an additional high-dimensional parameter: the cost of information acquisition. The appropriate values for this parameter have been the subject of intense inquiry. Some studies, such as Caplin and Dean (2015), de Oliveira et al. (2017), Dean and Neligh (2019), Dewan and Neligh (2020), and Caplin et al. (2020), seek to answer this question empirically, developing tools for identifying, measuring, and testing the agent's cost function in experiments. Other studies take an axiomatic approach, advocating for classes of parameterized cost functions based on their characterizing features (e.g., Pomatto, Strack, and Tamuz, 2020; Hébert and Woodford, 2021a; Caplin, Dean, and Leahy, 2021). An alternative string of papers explores which cost functions can be microfounded via dynamic learning (e.g., Morris and Strack, 2019; Bloedel and Zhong, 2020; Hébert and Woodford, 2021b) or as a physical cost of performing experiments (Denti et al., 2021). In this paper, we ask a complementary question: In what ways can one use the agent's information-acquisition cost to predict her behavior?

To answer our question, we study a canonical flexible-learning model in which an agent chooses from a finite set of alternatives, the benefits from which depend on a stochastic state. Before making her decision, the agent chooses what signal to acquire about this state. Learning comes at a cost, which the agent subtracts from the expected benefit she derives from her final decision. After choosing her information, the agent observes a signal realization and takes an action. We refer to the resulting probability in which the agent takes each action conditional on the state as the agent's *stochastic choice rule* (SCR).

We begin our analysis by asking whether knowing both the agent's costs and her benefits enables one to forecast the agent's behavior. Clearly, one's prediction should solve the agent's optimization problem and so must coincide with this solution whenever the latter is unique. By contrast, the appropriate forecast is unclear in the presence of multiple optima. Our results show, however, that such multiplicity is rare: the set of utilities that admit multiple optima is negligible, both topologically and measure theoretically. Hence, combined with the agent's benefits, one can typically use the agent's cost function to precisely predict the agent's chosen SCR.

Our next set of results pertain to one's ability to forecast the agent's choices using only her cost of information acquisition. Proposition 2 shows this exercise is essentially futile: holding the agent's costs fixed, one can approximate any finite-cost SCR arbitrarily well with SCRs that are optimal for some utility function. Therefore, other than ruling out some SCRs as technologically infeasible, the agent's learning costs make no predictions that could be falsified with finite data.

Proposition 2 leaves open the possibility that the approximating SCRs are rationalized via indifference or with a small set of utilities. For example, if learning is free, the agent optimally takes all actions with positive probability in all states only if her action does not influence her payoffs. In such cases, an analyst may be able to obtain meaningful predictions by ruling out a knife-edge set of utilities or by using other considerations to refine the set of optimal SCRs. Theorem 1 shows such refinements cannot meaningfully increase the set of restrictions imposed on the agent's behavior by her cost function, provided that this function is strictly increasing in informativeness and the set of feasible information structures is sufficiently flexible.

The above-mentioned results imply the agent's cost function on its own imposes few restrictions on the agent's behavior in a single menu. However, we show the agent's costs can significantly restrict the way the agent behaves *across* menus. To make this point, we focus on a novel class of smooth cost functions that we call *iteratively differentiable*. An advantage of such cost functions is that one can solve for the agent's optimal SCR given a fixed utility function using a first-order condition (Theorem 2). One can also apply this condition to do the reverse, namely, to solve for the utility functions that make a fixed SCR optimal. In Theorem 3, we use the ability to invert utility from choices to show that, for finite-valued iteratively differentiable costs satisfying an infinite-slope condition, one can often use the agent's actions in one menu to pin down her behavior in all submenus. Thus, under the theorem's conditions, once the analyst knows the agent's cost of learning, the agent's choice when facing the grand menu is the sole remaining degree of freedom in the agent's behavior.

The current study delineates the kind of data needed for testing various hypotheses regarding the agent's information-acquisition costs. Such tests must rely either on variation in the agent's menu or on direct payoff information, and cannot be based solely on the agent's behavior in a single menu. Consistent with this observation, several studies consider the problem of testing whether a dataset consisting of choices and utilities over outcomes (where the latter is either given or inferred) from multiple decision problems is consistent with costly information acquisition, and if so, whether one can use these data to make inferences about the agent's cost of information (e.g., Caplin and Dean, 2015; de Oliveira et al., 2017; Caplin et al., 2020; Chambers, Liu, and Rehbeck, 2020; Dewan and Neligh, 2020; Dillenberger, Krishna, and Sadowski, 2020; Lin, 2022). By studying the opposite question, namely, how one can use cost information to predict the agent's behavior, our paper shows one cannot hope for a more parsimonious class of tests than those considered by this literature.

A few other studies take an axiomatic approach to pinning down the agent's cost function. Some papers, such as Caplin, Dean, and Leahy (2017), de Oliveira (2014), and Denti (2022), study the properties that are necessary and sufficient for the agent's behavior to come from a cost function of a given class. Others studies, such as Mensch (2018), Pomatto, Strack, and Tamuz (2020), and Hébert and Woodford (2021a), characterize classes of cost functions by stating properties directly on the cost function itself rather than on the behavior it generates.[1] Several related papers such as Morris and Strack (2019), Bloedel and Zhong (2020), and Hébert and Woodford (2021b) ask which static cost functions can be micro-founded as coming from a dynamic learning process. Recently, Denti et al. (2021) obtain conditions under which a cost function defined over the agent's distribution of posterior beliefs over all priors can be micro-founded as a prior-independent cost over experiments. Our work differs from these studies in that we do not ask whether a certain cost function is reasonable. Instead, we consider the class of restrictions imposed by a fixed cost function on the agent's behavior.

2. Model

An agent makes a decision from a finite set A of actions with $|A| > 1$. The payoff from each action depends on a payoff state ω belonging to some compact metric space Ω and distributed according to some probability measure $\mu_0 \in \Delta\Omega$.[2] Without loss of generality, we take μ_0 to have full support. The agent's utility from choosing action $a \in A$ in state $\omega \in \Omega$ is $u_a(\omega)$. We assume $u_a \in L^1(\mu_0)$, meaning the agent's expected payoff $\mathbb{E}[u_a]$ from every action $a \in A$ is well defined, and refer to $u := (u_a)_{a\in A} \in \mathcal{U} := L^1(\mu_0)^A$ as the agent's **utility function**.[3]

[1]Relatedly, Frankel and Kamenica (2019) study axiomatically the appropriate measures of the ex-post value of information.

[2]We view any Polish space Y as a measurable space, endowed with its Borel sigma-algebra. Let ΔY denote the set of probability measures on this measurable space. Unless otherwise stated, we endow ΔY with the weak* topology generated by continuous bounded functions, which in turn makes ΔY a compact metrizable space if Y is compact. We let $\operatorname{supp}(\gamma)$ denote the support of γ for any $\gamma \in \Delta Y$.

[3]For any random variable $f : \Omega \to \mathbb{R}$, let $\mathbb{E}[f] := \int f(\omega)\, \mu_0(\mathrm{d}\omega)$ whenever the latter Lebesgue integral is well defined. Recall the space $L^1(\mu_0)$ is the set of all measurable functions $f : \Omega \to \mathbb{R}$ such that $\mathbb{E}|f| < \infty$, modulo μ_0-almost sure equivalence, equipped with the $L^1(\mu_0)$ norm, $\|f\|_1 := \mathbb{E}|f|$.

Information Policies. Before taking her action, our agent chooses what signal to observe in order to learn about ω. We assume our agent is Bayesian, and model the agent's information via the distribution of her posterior beliefs, $p \in \Delta\Delta\Omega$. Specifically, we allow the agent to choose any p whose average equals the prior, $\int \mu\, p(\mathrm{d}\mu) = \mu_0$, which is equivalent to letting her select any p that originates from some signal structure.[4] We refer to any p that averages back to the prior as an **information policy**, and denote the set of all information policies by $\mathcal{P}$. Occasionally, we will refer to the set $\mathcal{P}^F \subseteq \mathcal{P}$ of **simple information policies**, namely those with finite support. Relatedly, a **simply drawn posterior** is a posterior that belongs to the support of some simple information policy.[5]

We order information policies via informativeness in the sense of Blackwell (1953). Specifically, we say $p \in \mathcal{P}$ is **more informative** than $q \in \mathcal{P}$ (written as $p \succeq q$) if p is a mean-preserving spread of q.[6] An information policy p is **strictly more informative** than q (written $p \succ q$) if $p \succeq q$ and $p \neq q$. Intuitively, p is more informative than q if observing p is equivalent to observing q and an additional signal. For a review of the connection between this information ranking and other notions of informativeness, see Khan et al. (2019).

Information Costs. Information comes at a cost that is summarized via a function taking values in the extended reals,

$$C : \mathcal{P} \to \overline{\mathbb{R}} := \mathbb{R} \cup \{\infty\}.$$

By letting C take a value of infinity, we allow it to encode constraints on the set of feasible information policies. We assume C is a lower-semicontinuous convex function that is **proper**—that is, not globally equal to ∞—and let $\mathcal{P}^C = C^{-1}(\mathbb{R})$ denote its **(effective) domain**, namely, the nonempty set of information policies that can be induced at finite cost. We also require C to be **monotone**, meaning $C(p) \geq C(q)$ whenever $p \succeq q$. Sometimes, we focus on costs that are **strictly monotone**, by which we mean $C(p) > C(q)$ holds whenever $p \succ q$ and $q \in \mathcal{P}^C$. As we explain in section 7, assuming C is convex and monotone is without loss of generality.

[4]See, for example, Aumann and Maschler (1995), Kamenica and Gentzkow (2011), or Benoît and Dubra (2011). For the sake of completeness, Appendix C.2 provides a proof that applies to the case of infinite states.

[5]Equivalently, simply drawn posteriors are the elements of $\{\mu \in \Delta\Omega : \ \epsilon\mu \leq \mu_0 \text{ for some } \epsilon > 0\}$.

[6]That is, some measurable $m : \Delta\Omega \to \Delta\Delta\Omega$ exists such that $\int \tilde{\mu}\, m(\mathrm{d}\tilde{\mu}|\mu) = \mu$ for all $\mu \in \Delta\Omega$, and $p(D) = \int m(D|\mu)\, q(\mathrm{d}\mu)$ for all Borel $D \subseteq \Delta\Omega$.

The Agent's Problem. After choosing her information, the agent observes her signal realization and takes an action. An **action strategy** is a measurable mapping, $\alpha : \Delta\Omega \to \Delta A$, where $\alpha(a|\mu)$ is the probability the agent chooses action a when her posterior belief is μ. A **strategy** is an information policy p paired with an action strategy α. The agent's payoff from the strategy (p, α), given benefit $u \in \mathcal{U}$, is

$$\int_{\Delta\Omega} \int_A u_a \, \alpha(\mathrm{d}a|\mu) \, p(\mathrm{d}\mu) - C(p).$$

Observe some strategy yields a finite value, because some information policy yields finite cost. We say (p, α) is u**-optimal** if it maximizes the above objective among all possible strategies.

Stochastic Choice Rules. We summarize the agent's behavior via a vector of (essentially bounded) mappings,

$$s = (s_a)_{a \in A},$$

where $s_a \in L^\infty(\mu_0)$ gives the conditional probability the agent takes action a given the state.[7] Because probabilities are positive and add up to 1, we must have $s_a \geq 0$ for all a and $\sum_{a \in A} s_a = \mathbf{1}$. We refer to every $s \in \mathcal{S} := L^\infty(\mu_0)^A$ that satisfies these constraints as a **stochastic choice rule (SCR)**,[8] and denote the set of all SCRs by S.

The support of an SCR s is the set of actions it generates with positive probability, $\operatorname{supp} s = \{a \in A : s_a \neq \mathbf{0}\}$. An SCR s has **full support** if it uses all actions, that is, if $\operatorname{supp} s = A$. The SCR has **conditionally full support** if it uses all actions in all states, meaning s_a is μ_0-almost surely strictly positive for every $a \in A$.

A strategy (p, α) **induces** an SCR s if for every action $a \in A$ and event $\hat{\Omega} \subseteq \Omega$,

$$\int \alpha(a|\mu) \, \mu(\hat{\Omega}) \, p(\mathrm{d}\mu) = \mathbb{E}\left[\mathbf{1}_{\hat{\Omega}} \, s_a\right].$$

An information policy $p \in \mathcal{P}$ **can induce** $s \in S$ if p can describe the information the agent receives in some strategy that results in s, that is, if (p, α) induces s for some α.

[7] Recall $L^\infty(\mu_0)$ is the space of all measurable functions $f : \Omega \to \mathbb{R}$ (identified up to μ_0-almost sure equality) that are bounded μ_0-almost surely, equipped with the $L^\infty(\mu_0)$ norm, $\|f\|_\infty := \operatorname{ess\,sup} |f|$.

[8] Because conditional probabilities are well defined only up to almost-sure equivalence, stochastic choice rule naturally lives in $\mathcal{S} = L^\infty(\mu_0)^A$, which does not distinguish between functions that agree μ_0-almost surely. An equivalent formalism (given a disintegration theorem) would have stochastic choice rules living in the set of probability measures over $A \times \Omega$ with marginal μ_0 on Ω.

Rationalizability We are interested in understanding which stochastic choice rules are optimal for a *given* objective, which are optimal for *some* objective, and which can be *uniquely* optimal. Given a utility function $u \in \mathcal{U}$, we say s is u-**rationalizable** if it is induced by some u-optimal strategy (p, α). A stochastic choice rule is **uniquely** u-**rationalizable** if it is the only u-rationalizable SCR. An SCR that is u-rationalizable for *some* u is **rationalizable**. We also say s is **uniquely rationalizable** if it is uniquely u-rationalizable for some u.[9]

Example Cost Functions

In this section, we provide some example cost functions.

Example 1 (Mutual Information Costs)**.** *Let $K : \Delta\Omega \to \overline{\mathbb{R}}$ be the Kullback-Leibler divergence from the prior μ_0,*

$$K(\mu) = \begin{cases} \int \log \frac{\mathrm{d}\mu}{\mathrm{d}\mu_0}(\omega)\, \mu(\mathrm{d}\omega) & \text{if } \mu \ll \mu_0, \\ \infty & \text{otherwise.} \end{cases}$$

*By Posner (1975), K is lower semicontinuous. The **mutual information** cost function is given by*

$$C(p) = \int K(\mu)\, p(\mathrm{d}\mu).$$

This cost function was first introduced by Sims (1998; 2003; 2006) and has served as the workhorse cost function of the literature on rational inattention. See Csiszár (1974), Matějka and McKay (2015), Caplin, Dean, and Leahy (2019), and Denti, Marinacci, and Montrucchio (2020) for characterizations of optimal behavior under mutual information costs.

Example 2 (Posterior Separable Costs)**.** *Posterior separable costs are a generalization of mutual information costs introduced by Caplin, Dean, and Leahy (2017). Let $\mathcal{C}$ be the set of convex, lower semicontinuous functions from $\Delta\Omega$ to $\overline{\mathbb{R}}$ that assign μ_0 a finite value.*[10,11] *We say C is **posterior separable** if some $c \in \mathcal{C}$ exists such that every $p \in \mathcal{P}$ has*

$$C(p) = \int c(\mu)\, p(\mathrm{d}\mu).$$

[9] In other words, s is uniquely rationalizable if some u exists such that s is the only u-rationalizable SCR; this notion is distinct from requiring that u be the unique utility that rationalizes s.

[10] Given the other assumptions on elements of $\mathcal{C}$, the assumption that $c(\mu_0)$ is finite is equivalent to the induced information cost C being proper.

[11] When Ω is finite and c is globally finite, c is lower semicontinuous if and only if it is continuous.

In addition to mutual information, the class of posterior separable costs includes the log-likelihood ratio cost function axiomatized by Pomatto, Strack, and Tamuz (2020) and the neighborhood-based cost function studied by Hébert and Woodford (2021a).

Example 3 (Transformed Costs)**.** *Let $\psi : \mathbb{R} \to \overline{\mathbb{R}}$ be a nondecreasing, proper, convex, lower semicontinuous function. Then,*

$$C(p) = \psi\left(\int c(\mu)\, p(\mathrm{d}\mu)\right)$$

satisfies our assumptions for any $c \in \mathcal{C}$.

Example 4 (Quadratic Costs)**.** *Let $\tilde{c} : \Delta\Omega \times \Delta\Omega \to \mathbb{R}$ be a symmetric, lower semicontinuous function that is convex in each argument. Let*

$$C(p) = \int\int \tilde{c}(\tilde{\mu}, \mu)\, p(\mathrm{d}\tilde{\mu})\, p(\mathrm{d}\mu).$$

Then, C satisfies our assumptions as long as $\tilde{c}$ is positive semidefinite*, namely, as long as*

$$\int\int \tilde{c}(\tilde{\mu}, \mu)\, (p - q)(\mathrm{d}\tilde{\mu})\, (p - q)(\mathrm{d}\mu) \geq 0$$

holds for all $p, q \in \mathcal{P}$.

Example 5 (Maximum over a Set)**.** *Let $\tilde{\mathcal{C}} \subseteq \mathcal{C}$ be a compact (with respect to the supremum norm) set of continuous functions. Then,*

$$C(p) = \max_{c \in \tilde{\mathcal{C}}} \int c(\mu)\, p(\mathrm{d}\mu)$$

satisfies our assumptions.

3. Cost Minimization

We begin our analysis by studying the cheapest way to induce a given SCR. Specifically, we solve

$$\kappa(s) = \inf_{p \in \mathcal{P}} C(p) \text{ s.t. } p \text{ can induce } s.$$

Note s can be rationalizable only if the above program admits a solution: otherwise, no (p, α) that induces s can ever be optimal, because one can always attain the same utility with lower costs.

To solve the cost-minimization program, observe that every s can be viewed as a signal structure whose realizations take the form of recommended actions. Therefore, one can apply Bayes rule to transform any s into its associated information policy, $p^s \in \mathcal{P}$. Formally, let

$$p_a^s := \mathbb{E}\left[s_a\right]$$

be the ex-ante probability that s generates the recommendation a. Whenever $p_a^s > 0$, Bayes' rule dictates the agent's posterior belief μ_a^s conditional on the realized action recommendation being a is given by[12]

$$\mu_a^s(\mathrm{d}\omega) = \frac{s_a(\omega)}{p_a^s}\mu_0(\mathrm{d}\omega).$$

It follows one can write p^s as

$$p^s = \sum_{a \in A} p_a^s \delta_{\mu_a^s},$$

where $\delta_{\mu_a^s} \in \Delta\Delta\Omega$ denoes the distribution that generates μ_a^s with probability 1. We follow the literature (Caplin and Dean, 2015) and refer to μ_a^s as the **revealed posterior** of a given s, and p^s as s's **revealed information policy**.

Lemma 1. *Policy $p \in \mathcal{P}$ can induce $s \in S$ if and only if $p \succeq p^s$. Therefore, $\kappa(s) = C(p^s)$.*

Similar results are prevalent in the literature for less general settings (e.g., Caplin and Dean, 2015). Despite the difference in generality, the intuition is identical. If (p, α) induces s, one can generate p^s by first drawing a posterior-action pair (μ, a) according to (p, α) and then revealing only the realized action to the agent. Clearly, seeing a alone is less informative than seeing both a and μ. But because a is independent of the state conditional on μ, seeing μ and a together is just as informative as observing μ on its own. In other words, p is more informative than p^s. Thus, the lowest-cost way of inducing s is given by p^s.

Toward understanding the structure of the agent's indirect cost function κ, we now establish a few facts about the connection between stochastic choice rules and their revealed information policies.

Lemma 2. *The following hold:*

(i) If $(s^n)_{n=1}^\infty$ weak converges to s, then $(p^{s^n})_{n=1}^\infty$ converges to p^s.*[13]

[12] When $p_a^s = 0$, we can let $\mu_a^s \in \Delta\Omega$ be arbitrary wherever the term appears.

[13] A sequence $(s^n)_{n=1}^\infty$ weak* converges to s if $\mathbb{E}\left[u_a s_a^n\right]$ converges to $\mathbb{E}\left[u_a s_a\right]$ for all $u \in \mathcal{U}$ and $a \in A$. Note the weak* topology on $S \subset \mathcal{S}$ is determined by its convergent sequences because the predual $\mathcal{U}$ is separable.

(ii) For any s, t in S and $f \in (0,1)$,

$$(1-\beta)p^s + \beta p^t \succeq p^{(1-\beta)s+\beta t}. \tag{1}$$

(iii) Moreover, the information ranking in (1) is strict whenever some $a \in \mathrm{supp}(s) \cap \mathrm{supp}(t)$ exists such that $\mu_a^s \neq \mu_a^t$.

Property (i) says weak* convergence of SCRs implies convergence of their induced information policies. Denti et al. (2021) prove a few related continuity properties under different topological assumptions. They also establish a version of property (ii) that applies to settings with infinite actions but finite states. This property shows the information policy revealed by a convex combination of SCRs is less informative than the convex combination of the two revealed policies.

We now explain Part (iii), which is novel and plays a key role in the sequel. Intuitively, both $(1-\beta)p^s + \beta p^t$ and $p^{(1-\beta)s+\beta t}$ come from a signal structure that randomly determines which of s and t to use to generate the agent's signal realization. However, whereas $p^{(1-\beta)s+\beta t}$ is obtained by showing the agent only the realized signal, to obtain $(1-\beta)p^s+\beta p^t$, one must also reveal to the agent whether this realization came from s or t. Hence, one can generate $(1-\beta)p^s + \beta p^t$ by using $p^{(1-\beta)s+\beta t}$ to draw a posterior $\mu_a^{(1-\beta)s+\beta t}$, and then splitting this posterior in a mean-preserving way between μ_a^s and μ_a^t. Whenever μ_a^s and μ_a^t differ, this split strictly changes the agent's beliefs and so must come from the agent learning more information.

Given the above lemma, one can immediately deduce that κ must satisfy two extemely useful properties. First, κ is weak* lower semicontinuous: if $(s^n)_{n=1}^\infty$ converges to s, then $(p^{s^n})_{n=1}^\infty$ converges to p^s, so that

$$\liminf_{n\to\infty} \kappa(s^n) = \liminf_{n\to\infty} C(p^{s^n}) \geq C(p^s) = \kappa(s),$$

where the inequality follows from lower semicontinuity of C. Second, κ is convex, because every $s, t \in S$ and every $\beta \in (0,1)$ satisfy

$$\begin{aligned} \kappa((1-\beta)s + \beta t) &= C(p^{(1-\beta)s+\beta t}) \\ &\leq C((1-\beta)p^s + \beta p^t) \\ &\leq (1-\beta)C(p^s) + \beta C(p^t) = (1-\beta)\kappa(p^s) + \beta\kappa(p^t), \end{aligned} \tag{2}$$

where monotonicity of C implies the first inequality, and convexity of C delivers the second.

Thus, we have obtained the following corollary.

Corollary 1. *The indirect cost κ is proper, convex, and weak* lower semicontinuous.*

4. Knowing Both Costs and Benefits

In this section, we assume knowledge of both the agent's information acquisition costs C and her payoffs from taking different actions, u. We show such knowledge enables the analyst to precisely predict the agent's conditional action distribution, except in a knife-edge set of utilities. Toward this goal, we first rephrase the agent's problem so that it takes the agent's SCR s as its decision variable. Given u, the benefit from choosing s is given by

$$\mathbb{E}[u \cdot s] = \mathbb{E}\left[\sum_{a \in A} u_a s_a\right].$$

Because $\kappa(s)$ gives the minimum cost at which the agent obtains s, we get that s is u-rationalizable if and only if it solves the program

$$\max_{s \in S} \left[\mathbb{E}[u \cdot s] - \kappa(s)\right]. \tag{3}$$

Observe the above program always admits a solution: because κ is weak* lower semicontinuous, the program maximizes a weak*-upper-semicontinuous objective over a weak*-compact set.[14]

However, that the agent's program can be solved does not mean we can predict her behavior: whenever the program admits multiple optima, the agent's behavior remains undetermined. Our next result shows such multiplicity arises only by accident.

Proposition 1. *The set of utilities with multiple rationalizable SCRs is meager and shy.*[15]

For intuition, recasting the agent's maximization program in geometric terms is useful. To that end, recall s is u-rationalizable if and only if

$$\mathbb{E}[u \cdot s] - \kappa(s) \geq \mathbb{E}[u \cdot t] - \kappa(t) \text{ for all } t \in S,$$

[14] Compactness follows from the Banach-Aloaglu theorem (see, e.g., Theorem 6.21 in Aliprantis and Border, 2006).

[15] A set is **meager** if it is a countable union of nowhere dense sets. A set $\mathcal{V} \subseteq \mathcal{U}$ is **shy** if some probability measure $\nu \in \Delta\mathcal{U}$ with compact support assigns zero measure to every translation of $\mathcal{V}$, that is, if $\nu(\mathcal{V} + u) = 0$ for all $u \in \mathcal{U}$. Shy sets generalize Lebesgue null sets beyond the case of finite dimensions (see Hunt, Sauer, and Yorke, 1992). Thus, the proposition implies that for finite Ω, the set of utilities that admit multiple rationalizable SCRs is Lebesgue null.

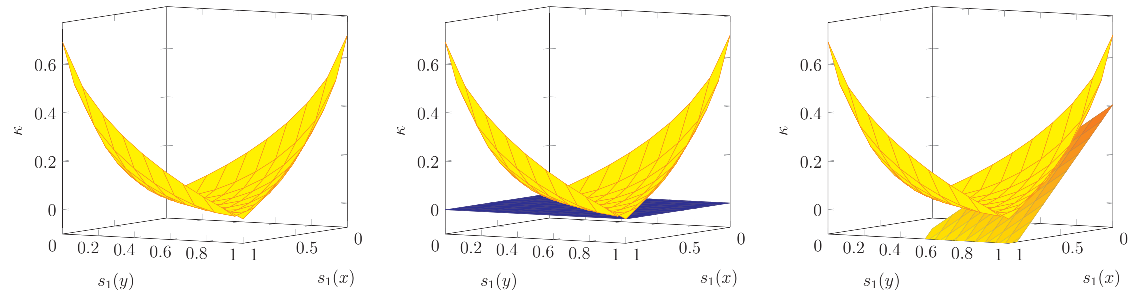

Figure 1: Graphs of the indirect cost function κ and some of its subgradients for the case in which $\Omega = \{x, y\}$, $A = \{0, 1\}$, and C is given by mutual information (see Example 1). The cost function is drawn as a function of $s_1(x)$ and $s_1(y)$.

which is clearly equivalent to

$$\kappa(t) \geq \kappa(s) + \mathbb{E}[u \cdot (t - s)] \text{ for all } t \in S.$$

The above condition has a geometric interpretation: s being u-rationalizable is equivalent to u being a subgradient of κ at s.

The relationship between subgradients and optimality is a standard fact from convex analysis (see, e.g., Theorem 23.5 in Rockafellar, 1970). Geometrically, u is a subgradient of κ at s if the function $t \mapsto \kappa(s) + \mathbb{E}[u \cdot (t - s)]$ defines a hyperplane whose graph supports the epigraph of κ at the point $(s, \kappa(s))$. Figure 1 visualizes this condition when $\Omega = \{x, y\}$, $A = \{0, 1\}$, and C is given by mutual information. In this case, each s can be summarized by the probability it takes action 1 in each state, $(s_1(x), s_1(y))$. The left panel illustrates κ as a function of these probabilities. When $s_1(x) = s_1(y)$, the action generated by s is uninformative about Ω, and so the agent's learning costs are minimized. Such SCRs are rationalizable by any utility that is constant across both actions. Figure 1's middle panel depicts the supporting hyperplane that results from such utility functions. As can be seen, the resulting hyperplane lies everywhere below the graph of κ, touching the graph only at the SCRs that are rationalizable via such utilities. For non-constant utilities, the agent usually finds it optimal to collect some information about the state. The figure's right panel depicts a supporting hyperplane that corresponds to such a situation.

Proposition 1 essentially follows from the fact that most hyperplanes that support the graph of a well-behaved convex function admit a unique support point. Figure 2 illustrates a sense in which this fact holds for a convex function ϕ defined over a closed subinterval of $\mathbb{R}$. In one dimension, a subgradient is defined via the slope of the corresponding line. For a line

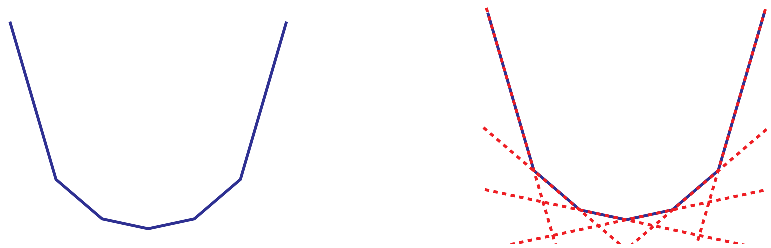

Figure 2: A convex function over an interval with the set of its subgradients that attain multiple support points.

to support ϕ in multiple points, the line's slope must equal the slope of ϕ in an interval over which ϕ is affine. Clearly, at most countably many such intervals can exist, and so the set of gradients that admit multiple support points must be countable as well. In fact, one can show this set is not only countable: it is also meager and Lebesgue null, two properties that generalize to higher dimensions.

5. Knowing Costs Only: An Impossibility Result

In this section, we study the analyst's ability to predict the agent's behavior without knowing anything about her utility. Thus, we take the agent's cost function as given and ask which SCRs can be optimal for some utility function, that is, which SCRs are rationalizable.

Clearly, the agent's cost function places some restrictions on the set of rationalizable SCRs. For instance, s can only be rationalizable if it is feasible; that is, only if it is in the set of SCRs that are attainable at a finite cost,

$$S^\kappa := \kappa^{-1}(\mathbb{R}).$$

The agent's cost function can also rule out some SCRs whose costs are finite. For example, whenever Ω is finite and C equals mutual information (Example 1), only SCRs that use all actions in their support in all states are rationalizable; that is, s is rationalizable only if s_a is strictly positive almost surely whenever it is not identical to zero (Caplin, Dean, and Leahy, 2019). Note, however, that this restriction has little practical relevance, because every SCR violating the restriction can be approximated arbitrarily well by SCRs that satisfy it.

The above discussion raises the following question: Does C impose any meaningful restrictions on the set of rationalizable SCRs beyond feasibility? The following result shows the answer is negative. To state the result, call a set of SCRs $\tilde{S}$ **uniformly dense** in S^κ if every $s \in S^\kappa$ and $\epsilon > 0$ admit some t in $\tilde{S}$ such that each $a \in A$ and μ_0-almost every $\omega \in \Omega$

have $|s_a(\omega) - t_a(\omega)| \le \epsilon$.

Proposition 2.

(i) The set of rationalizable SCRs is uniformly dense in S^κ.

(ii) If Ω is finite, every SCR in the relative interior of S^κ is rationalizable.

The proposition's argument is based on the above-mentioned geometric relationship between SCRs and their rationalizing utilities: because s being u-rationalizable is equivalent to u being a subgradient of κ at s, the SCR s is rationalizable if and only if the set of all such subgradients

$$\partial\kappa(s) := \left\{ u \in \mathcal{U} : \ \kappa(t) \ge \kappa(s) + \mathbb{E}[u \cdot (t - s)] \text{ for all } t \in S \right\},$$

also known as κ's **subdifferential** at s, is nonempty. The proposition then follows from classic results in convex analysis. For Part (i), we observe κ is the convex conjugate of a continuous convex function, and so one can apply a dual version of the Brøndsted-Rockafellar theorem (Brøndsted and Rockafellar, 1965). For Part (ii), we note that whenever Ω is finite, one can view S as a subset of finite dimensional Euclidean space, enabling us to appeal to results assuring subdifferentiability on the interior of a convex function's domain (e.g., Theorem 23.4 in Rockafellar, 1970).

Caplin, Dean, and Leahy (2021) prove a specialization of Part (ii) for the case in which C is posterior separable and when all SCRs in the interior of S have a finite cost. Their argument constructs a utility function that rationalizes an interior s by choosing u_a to be an appropriate member of c's subdifferential at μ_a^s. This construction suggests a connection between the subdifferentials of c and κ whenever C happens to be posterior separable. In section 7, we discuss this connection in more detail.

One potential concern with Proposition 2 is that it could be driven by indifference. For an extreme example, suppose learning is free; that is, $C(p) = 0$ for all p. In this case, only utility functions that do not depend on the agent's action (meaning $u_a = u_{a'}$ for all a, a') can rationalize a conditionally full-support SCR. Such indifference seems problematic, for two reasons. First, indifference allows a negligible set of utilities to rationalize many SCRs, allowing one to obtain substantive behavioral predictions by ruling out utilities that are in some sense rare. Second, indifference leaves open the possibility of obtaining meaningful restriction on choices via the introduction of an appropriate way to refine optimal behavior.

Next, we state some assumptions under which one can rationalize essentially every SCR without relying on indifference.

Assumption A1.

(i) The cost function C is strictly monotone.

(ii) The cardinality ranking $|\Omega| \geq |A|$ holds.

(iii) The domain S^κ has a nonempty (norm) interior in S.

Part (i) means learning more always comes at a strictly positive cost. This part rules out indifference driven by some information being free. Part (ii) requires the set of states to be richer than the set of actions. Whenever part (ii) fails, the information policy p^s associated with any full-support s can be written as the convex combination of two other information policies, each of which is associated with a different SCR. In other words, one can find $t, \tilde{t} \in S \setminus \{s\}$ and $\beta \in (0,1)$ such that $p^s = (1-\beta)p^t + \beta p^{\tilde{t}}$. Hence, if C is affine around p^s, s cannot possibly be uniquely rationalizable. The next result proves this assertion formally.

Claim 1. *Suppose $s \in S^\kappa$ is such that $|\text{supp}\, s| > |\Omega|$. If C is affine in a neighborhood of p^s, s is not uniquely rationalizable.*

An immediate consequence of the above claim is that no full support SCR is uniquely rationalizable when C is posterior separable and $|A| > |\Omega|$. Hence, Assumption A1(ii) is necessary for obtaining unique rationalizability in the posterior-separable case, which plays an important role in the literature. Assumption A1(iii) plays a similar role, by requiring the agent's information acquisition technology to be flexible in a local sense.[16] Without Assumption A1(iii), the cost C can restrict the agent to a low-dimensional set of information policies, a restriction with similar implications to violations of Assumption A1(ii). To illustrate, suppose the state is divided into two separate connected components, and the only finite-cost signals are ones that cannot distinguish between states within each component. For example, $\Omega = [0,1] \cup [2,3]$, and any SCR that is not constant on $[0,1]$ and on $[2,3]$ has infinite cost. Because such an environment is essentially equivalent to a binary-state model, one could obtain an analogue of Claim 1 whenever more than two actions have positive probability.

Our next goal is to show that, under Assumption A1, the inability to predict the agent's behavior using her information costs alone is not driven by indifference, nor does it rely on a

[16]This assumption holds, for example, if the agent can obtain a signal that reveals the state with probability ϵ, and provides no information otherwise. It also holds for mutual information costs (Example 1).

negligible set of utilities. This result is based on a particular set of SCRs that we call linearly independent. Formally, an $s \in S$ is **linearly independent** if $\{s_a\}_{a\in A} \subseteq L^\infty(\mu_0)$ consists of $|A|$ linearly independent elements.

Linearly independent SCRs possess the following useful property: given any linearly independent s and any other SCR $t \neq s$, some action a exists that reveals a different posterior under t than under s. Formally, some $a \in \text{supp}(s) \cap \text{supp}(t)$ exists such that $\mu_a^s \neq \mu_a^t$. This property follows from observing that an SCR is linearly independent if and only if it employs every action with positive probability and its revealed posteriors $\{\mu_a^s\}_{a\in A} \subseteq \Delta\Omega$ consist of $|A|$ affinely independent beliefs. Therefore, whenever s is linearly independent, one can split the prior across s's revealed posteriors in only one way, and so t reveals the same posterior for every action as s does only if the two SCRs are identical.

We now explain that κ is strictly convex through any linearly independent SCR whenever C is strictly monotone. The reason is that, by Lemma 2(iii), the property mentioned in the previous paragraph means information strictly decreases with convex combinations that involve linearly independent SCRs. Specifically, for any two SCRs $t \neq s$ such that s is linearly independent and every $\beta \in (0, 1)$,

$$(1-\beta)p^s + \beta p^t \succ p^{(1-\beta)s+\beta t}.$$

Hence, if C is strictly monotone, one can revisit the convexity argument for κ to obtain the following for all such s, t, and β:

$$(1-\beta)\kappa(s) + \beta\kappa(t) > \kappa((1-\beta)s + \beta t),$$

where the strict inequality follows from the fact that, in this case, the first inequality in chain (2) is strict. It follows that, for strictly monotone C, a linearly independent SCR is optimal only if it is uniquely so.

Proposition 3. *Suppose C is strictly monotone. If s is linearly independent, then κ is strictly convex through s.*[17] *Consequently, if s is u-rationalizable, it is uniquely u-rationalizable.*

Proposition 3 identifies the set of linearly independent SCRs as ones that cannot be rationalized using indifference when C is strictly monotone. This result generalizes a well-known condition for unique rationalizability under mutual information costs. Specifically, Caplin and Dean (2013) and Matějka and McKay (2015) show that, with mutual information costs,

[17] That is, κ is strictly convex on any line segment in S^κ that includes s.

s is uniquely u-rationalizable whenever the set $\left\{e^{u_a} : a \in A\right\}$ consists of $|A|$ affinely independent elements. To see why this result is a specialization of Proposition 3, recall that Matějka and McKay (2015) show that,[18] with mutual information costs, s is optimal when the utility is u only if

$$s_a(\omega) = \frac{p_a^s e^{u_a(\omega)}}{\sum_{b \in A} p_b^s e^{u_b(\omega)}}.$$

It is easy to verify that any s satisfying the above display equation must be linearly independent whenever $\left\{e^{u_a} : a \in A\right\}$ consists of $|A|$ affinely independent elements. Unique rationalizability then follows from Proposition 3.

Our next theorem complements the above observations by showing that under Assumption A1, most SCRs are linearly independent. It follows that the inability to predict the agent's behavior using her information costs alone is not driven by indifference. The theorem also shows this inability does not depend on a negligible set of utilities.

Theorem 1. *Suppose Assumption A1 holds. Then, a set of uniquely rationalizable SCRs exists that is uniformly dense in S^κ and is open if Ω is finite. Moreover, this set of SCRs is rationalized by an open set of utilities.*

In addition to showing learning costs alone impose no substantial restrictions on behavior, the above analysis also highlights the futility of searching for cost functions that require the agent's behavior to satisfy certain desiderata. To illustrate, consider the following property, suggested by Morris and Yang (2021): a cost function C satisfies **continuous choice** if only continuous SCRs are rationalizable.[19] Morris and Yang (2021) suggest continuous choice as a way to model agents who have a hard time distinguishing between similar states. Below, we show the only way to guarantee this property is to make discontinuous information structures infeasible. To show this result, we observe that continuity of an SCR is a closed property, and so can be satisfied by a dense subset of S^κ only if it satisfied by all of S^κ's elements.

Proposition 4.

(i) Every rationalizable SCR is continuous if and only if every SCR in S^κ is continuous.

(ii) If the domain S^κ has a nonempty (norm) interior in S, then every rationalizable SCR is continuous if and only if Ω is finite.

[18] See also Csiszár (1974).

[19] An SCR s is **continuous** if it admits a continuous version; that is, every a admits a continuous function $f_a : \Omega \to [0, 1]$ such that $f_a = s_a$ almost surely.

(iii) If Assumption A1 holds, then every uniquely rationalizable SCR is continuous if and only if Ω is finite.

Thus, our results suggest that except in cases where continuous choice is vacuous (i.e., when Ω is finite), it can only hold if discontinuous SCRs are infeasible, a property referred to by Morris and Yang (2021) as *infeasible perfect discrimination*.[20] In terms of C, this property is equivalent to C assigning infinite cost to every simple information policy that generates a μ with a discontinuous Radon-Nikodym derivative with respect to μ_0. In section 7, we note one can circumvent the need for infinite costs by requiring continuous choice to hold only for a restricted set of objectives. For example, one may require the agent's choice to be continuous for any bounded objective. As we explain later, such a requirement is equivalent to a slight weakening of Morris and Yang's (2021) *expensive perfect discrimination* property.

6. Cross-Menu Restrictions

The previous section shows learning costs alone impose no meaningful restrictions on the agent's behavior given a fixed menu. In this section, we highlight these costs do constrain how the agent behaves *across* menus. The reason is that the agent's actions in one menu allows the analyst to use the agent's cost function to extract information about her preferences. In fact, we show the information revealed by the agent's choices is particularly sharp when her cost function is smooth in a manner we make precise below. This sharpness allows us to prove the following result: whenever our smoothness conditions holds, most choices in one menu pin down the agent's actions in all submenus.

6.1. Smooth Costs

We introduce our smoothness condition in stages. We begin by defining a differentiability notion for C. To that end, we recall some standard notation. Given a convex set X in a real vector space, a convex function $f : X \to \overline{\mathbb{R}}$, and $x \in f^{-1}(\mathbb{R})$, define the **directional derivative** of f at x as $d_x^+ f : X \to \mathbb{R} \cup \{\pm\infty\}$ via

$$d_x^+ f(x') := \lim_{\epsilon \searrow 0} \tfrac{1}{\epsilon} \left[f(x + \epsilon(x' - x)) - f(x) \right].$$

[20] We should highlight that we are abusing terminology slightly: Morris and Yang (2021) require Ω to be a real interval, and say that a cost function satisfies infeasible perfect discrimination if it assigns finite cost only to SCRs that are *absolutely* continuous.

Say $c \in \mathcal{C}$ is **a derivative** of C at $p \in \mathcal{P}^C$ if $\int c(\mu)\, p(\mathrm{d}\mu)$ is finite and if every $p' \in \mathcal{P}^C$ has

$$d_p^+ C(p') = \int c(\mu)\, (p' - p)(\mathrm{d}\mu).$$

The cost function C is **differentiable at** $p \in \mathcal{P}^C$ if it admits a derivative at p.[21] We omit the dependency on the information policy and say C is **differentiable** whenever it is differentiable at all $p \in \mathcal{P}^C$.[22]

Intuitively, a cost function is differentiable if it prices small shifts in its information in a posterior-separable manner. Indeed, observe that the function

$$C_c(q) = \int c(\mu)\, q(\mathrm{d}\mu)$$

defines a posterior-separable cost function. In fact, this cost equals C whenever the latter is already posterior separable: in this case, C has a common derivative at all information policies.

In general, the derivative of a differentiable cost function that is not posterior separable depends on p. For a demonstration, consider the cost function in Example 3, $C(p) = \psi\left(\int c(\mu)\, p(\mathrm{d}\mu)\right)$. Whenever ψ is differentiable, this cost function admits

$$\psi'\left(\int c(\mu)\, p(\mathrm{d}\mu)\right) c(\cdot)$$

as its derivative at p, which depends on p whenever ψ is not affine. For another example, consider the cost function from Example 4. This cost function is differentiable, with a derivative at $p \in \mathcal{P}^C$ given by $2 \int \tilde{c}(\cdot, \mu)\, p(\mathrm{d}\mu)$. Clearly, this derivative typically changes as p varies.[23]

Our next result tightens the connection between posterior separability and differentiability. To state this result, define the indirect cost function associated with C_c,

$$\kappa_c(t) := C_c(p^t).$$

[21] Our definition of differentiability requires the derivative to be convex (because $c \in \mathcal{C}$). In the appendix, we show one can omit this requirement whenever c is finite and continuous.

[22] Our notion of differentiability is commonly used in the decision theory literature, where it is often called *Gâteaux differentiability* (e.g., Hong, Karni, and Safra, 1987; Cerreia-Vioglio, Maccheroni, and Marinacci, 2017). The definition is slightly different than the way Gâteaux differentiability is defined in convex analysis: whereas the latter requires the convergence to occur from all possible directions (e.g., Phelps, 2009; Borwein and Vanderwerff, 2010), we require convergence only from directions within the domain of C.

[23] For a non-differentiable cost function, consider Example 5 when $\tilde{\mathcal{C}}$ is finite and not equivalent to a singleton. In this case, one obtains a point of non-differentiability at any p on the boundary between two regions where the set of maximizers differs.

Theorem 2 below shows that, whenever C has full domain, s is u-rationalizable if and only if it is u-rationalizable when costs are given by the cost function's posterior separable approximation at p^s.

Theorem 2. *Fix some $s \in S$ and $u \in \mathcal{U}$. Suppose C is finite on $\mathcal{P}^F$ and that c is a derivative of C at p^s.*[24] *Then,*

$$s \in \text{argmax}_{t \in S} \left[\mathbb{E}\left[u \cdot t\right] - \kappa(t) \right]$$

if and only if

$$s \in \text{argmax}_{t \in S} \left[\mathbb{E}\left[u \cdot t\right] - \kappa_c(t) \right].$$

Note the above result does not tell us κ and κ_c have the same set of maximizers. The reason is that the cost function's derivative depends on the SCR around which the cost is approximated. In other words, u-rationalizability of s is equivalent to s being u-rationalizable under C_c only if c is a derivative of C at p^s. Unless C is posterior separable (in which case all SCRs admit a common derivative), different SCRs usually admit different derivatives.

Next, we define a differentiability notion for derivatives of C. Given $c \in \mathcal{C}$ and a simply drawn μ, we say $\nabla c_\mu \in L^1(\mu_0)$ is a **derivative** of c at μ if $\int \nabla c_\mu(\omega)\ \mu(\mathrm{d}\omega) = c(\mu)$, and every simply drawn $\mu' \in \Delta\Omega$ has

$$d^+_\mu c(\mu') = \int \nabla c_\mu(\omega)\ (\mu' - \mu)(\mathrm{d}\omega).$$

The function c is **differentiable** at μ if it admits a derivative there. Thus, c is differentiable if it can be locally approximated by an affine function.

Our key notion of smoothness requires C to admit a differentiable derivative. Specifically, we say C is **iteratively differentiable** at $p \in \mathcal{P}^F$ if it admits a derivative c at p that is differentiable at every $\mu \in \text{supp}\ p$. Thus, an iteratively differentiable cost is locally similar to a smooth, posterior-separable cost function.

The benefit of having iteratively differentiable costs is summarized in the following proposition, which tightly characterizes when an interior SCR is optimal.

Proposition 5. *Suppose s has full support, and C is finite on $\mathcal{P}^F$ and iteratively differentiable at p^s with derivative c. The following are equivalent for $u \in \mathcal{U}$:*

(i) SCR s is u-rationalizable; that is, $s \in \text{argmax}_{t \in S} \left[\mathbb{E}\left[u \cdot t\right] - \kappa(t) \right]$.

[24] One can weaken the requirement that C is finite on $\mathcal{P}^F$, assuming only that C is finite for simple information policies around p^s. See the appendix for more details. Proposition 5 below can be similarly strengthened.

(ii) Some $\lambda \in L^1(\mu_0)$ and $\gamma \in L^1(\mu_0)_+^A$ exist such that every $a \in A$ have

$$u_a = \lambda - \gamma_a + \nabla c_{\mu_a^s} \text{ and } \gamma_a s_a = 0.$$

One can view the proposition as establishing a Lagrange multiplier result for iteratively differentiable costs, with $\nabla c_{\mu_a^s}$ serving the role of the derivative of κ. To better see this interpretation, suppose Ω is finite and that κ is differentiable. In this case, s being u-rationalizable is equivalent to s solving the program

$$\begin{aligned} \max_{t \in \mathbb{R}^{A \times \Omega}} \quad & \Big[\mathbb{E}\left[u \cdot t\right] - \kappa(t)\Big] \\ \text{s.t.} \quad & t_a(\omega) \geq 0 \qquad \forall a \in A,\ \forall \omega \in \Omega, \\ & \sum_a t_a(\omega) = 1 \qquad \forall \omega \in \Omega. \end{aligned}$$

Applying a standard Lagrangian result (e.g., Pourciau, 1983) gives a multiplier $\gamma_a(\omega)$ for every instance of the first constraint and a multiplier $\lambda(\omega)$ for every instance of the second constraint such that s is optimal if and only if

$$u_a(\omega) - \frac{1}{\mu_0(\omega)} \frac{\partial \kappa}{\partial s_a(\omega)} + \gamma_a(\omega) - \lambda(\omega) = 0.$$

Observe the above display equation specializes to condition (ii) of the proposition, but with $\frac{1}{\mu_0(\omega)} \frac{\partial \kappa}{\partial s_a(\omega)}$ replacing $\nabla c_{\mu_a^s}$.

In addition to characterizing optimality of a full-support s, Proposition 5 also enables one to recover the agent's utility function from their behavior, up to a nuisance term. For an explanation, suppose s has conditionally full support and that C is iteratively differentiable at p^s with derivative c. Let u^s be the utility function defined via $u_a^s := \nabla c_{\mu_a^s}$ for all a. Then, Proposition 5 implies a utility u rationalizes s if and only if u generates the same optima for all menus. The reason is that u equals u^s plus an action-independent shift, λ, the addition of which has no impact on the set of maximizers.[25] The next section uses this observation to identify conditions under which the agent's choices in one menu are sufficient for pinning down her behavior for all submenus.

[25] Note $\gamma_a s_a = 0$ means $\gamma_a = 0$ when s has conditionally full support.

6.2. Unique Subset Predictions

We now use the results of the previous section to show the agent's learning costs restrict her behavior across menus. To model these restrictions, let $\mathcal{A}$ be all nonempty subsets of A. For any $B \in \mathcal{A}$, let $S_B = \{s \in S : \sum_{a \in B} s_a = \mathbf{1}\}$ be the set of SCRs that only use actions in B. Given a utility function u, we say $s \in S_B$ is **u-rationalizable over** B if s solves the agent's problem when the agent is restricted to only using actions in B; that is,

$$s \in \text{argmax}_{t \in S_B} \left[\mathbb{E}\left[u \cdot t\right] - \kappa(t)\right].$$

Say $s \in S$ **yields unique subset predictions** if every $B \in \mathcal{A}$ has a unique $t \in S_B$ that admits a utility that both rationalizes t over B and rationalizes s (over A). Thus, if the agent chooses from A according to s, and s yields unique subset predictions, we can deduce exactly what must be the agent's chosen SCR when she's restricted to B.

Next, we introduce assumptions that enable us to conclude many of the agent's behaviors in one menu uniquely pin down her choices in all submenus. To state the assumption, say an information policy is **fully mixed** if all posteriors in its support have a strictly positive Radon-Nikodym derivative with respect to μ_0.

Assumption A2.

(i) C is finite at every simple information policy.[26]

(ii) C is iteratively differentiable at every simple and fully mixed information policy.

(iii) If $p \in \mathcal{P}^F$ is not fully mixed, $d_p^+ C(\delta_{\mu_0}) = -\infty$.

The first two parts of the assumption require C to be smooth, as explained in the previous section. The third part requires C to have infinite slopes at simple information policies that are not fully mixed. This property, which is satisfied by mutual-information costs (Example 1), implies an SCR can be rationalizable only if it uses all actions in its support in all states. In particular, a full-support SCR is rationalizable only if it has conditionally full support. Therefore, Assumption A2 implies one can use Proposition 5 to recover the agent's utility function, up to a choice-irrelevant nuisance term, whenever she employs a full-support SCR. Our next result uses this fact to demonstrate that the agent's learning costs often impose strong restrictions on the agent's behavior across menus.

[26]This condition implies $S^\kappa = S$, and, in particular, implies A1(iii).

Theorem 3. *Under A1 and A2, SCRs yielding unique subset predictions are weak* dense.*

To prove the theorem, we first use Proposition 1 to find a dense set $\mathcal{U}_{\mathcal{A}}$ of utilities that attain a unique optimum at every menu. Using a continuity property of the subdifferential, we then show one can approximate any uniquely rationalizable s with SCRs that are rationalizable by some utility in $\mathcal{U}_{\mathcal{A}}$. Moreover, whenever s has conditional full support, the approximating SCRs can be taken to have the same property.[27] Focusing on one of the approximating SCRs, Proposition 5 implies all utilities that rationalize it generate the same optima for all menus. Because one of these rationalizing utilities comes from $\mathcal{U}_{\mathcal{A}}$, it follows that, for any given menu, all of these utilities uniquely rationalize the same SCR. In other words, each of the SCRs approximating s yields unique subset predictions. All that is left is to note Theorem 1 implies the set of uniquely rationalizable SCRs with full support (which in fact have conditionally full support due to the infinite-slope condition) is dense in S^κ.

7. Discussion

In this section, we discuss our model's assumptions, additional results, and the relationship to existing literature.

Partial Knowledge of Costs. Our model assumes the analyst knows the agent's learning costs exactly. Maintaining this stylized assumption enables us to study the degree to which the agent's learning costs pin down her behavior. In practice, many analysts may not know C exactly, but are instead capable of restricting it to belong to some set $\mathfrak{C}$. Some of our results also speak to this case. For example, Proposition 2 implies the set of SCRs that is rationalizable by some cost function in $\mathfrak{C}$ is uniformly dense in the set of SCRs that are feasible for some $C \in \mathfrak{C}$. Similarly, if each $C \in \mathfrak{C}$ satisfies A1, Theorem 1 immediately implies the set of uniquely rationalizable SCRs is uniformly dense in the set of SCRs that can be induced at finite C-cost for some $C \in \mathfrak{C}$. Proposition 1 also extends somewhat: if $\mathfrak{C}$ is countable, the set of utilities that does not generate a unique prediction for some $C \in \mathfrak{C}$ is meager and shy. Hence, for most utility functions, the analyst's uncertainty about the agent's behavior reduces to the uncertainty about C—provided $\mathfrak{C}$ is countable. The reason is that a countable union of meager and shy sets is itself meager and shy. By contrast, we do not know of an immediate way to extend Theorem 3 to accommodate multiple cost functions.

[27]This property of the approximating SCRs relies on Assumption A2(iii). We note this assumption is not needed when Ω is finite, because in this case, the set of conditionally full-support SCRs is open.

Unique Rationalizability and Strict Convexity. Our analysis showed that, under A1, one can rule out indifference as the source of the analyst's inability to predict behavior using the agent's learning costs. We focused on A1 because it accommodates the important case of posterior separable costs. An alternative way to obtain a similar result is to look at costs that are strictly convex. We now provide such a result.

Proposition 6. *Suppose C is strictly convex on $\mathcal{P}^C$, and $\mathcal{P}^C \neq \{\delta_{\mu_0}\}$. Then, a set of uniquely rationalizable SCRs exists that is uniformly dense in S^κ and is open if Ω is finite. Moreover, this set of SCRs is rationalized by an open set of utilities.*

As the above result highlights, some rationalizable SCRs need not be uniquely rationalizable even when C is strictly convex. The reason is that some convex combinations of SCRs change the way the agent randomizes over actions conditional on her information without changing the information itself. For example, suppose the action set is binary, $A = \{0, 1\}$, and take s and t to be the SCRs that respectively take action 1 and 0 regardless of the state. Clearly, both SCRs reveal an uninformative information policy, $p^s = p^t = \delta_{\mu_0}$. Moreover, the same is true for any convex combination of s and t, because any such combination results in the agent's actions being independent of the state. It follows the cost of any such convex combination is identical to the cost of s and t. In other words, even when C is strictly convex, κ is still affine over some line segments.

To prove Proposition 6, we show strict convexity of C implies κ is strictly convex over any line segment with an end point that satisfies the following property: every action is used with positive probability, and no two actions reveal the same posterior. Therefore, any rationalizable SCR with this property is uniquely rationalizable. The proposition's proof then proceeds as the proof of Theorem 1, but with SCRs with the previously mentioned property taking the role of the set of linearly independent SCRs.

Next, we note substituting strict convexity of C for A1 does not alter the conclusions of Theorem 3.

Proposition 7. *If C is strictly convex on its domain, $|\Omega| > 1$, and A2 holds, the set of SCRs that yield unique subset predictions is weak* dense.*

The argument for Proposition 7 follows the same reasoning as Theorem 3. We refer the reader to the appendix for the specific details.

Subdifferentials, Rationalizability, and Posterior Separable Costs. Among their many contributions, Caplin, Dean, and Leahy (2021) also identify a connection between rationalizability and subdifferentiability for the case in which C is posterior separable. More specifically, Caplin, Dean, and Leahy (2021) show that if Ω is finite and C is posterior separable, s is rationalizable if and only if c has a nonempty subdifferential at all beliefs in the support of s's revealed information policy; that is, $\partial c(\mu_a^s) \neq \varnothing$ for all $a \in \text{supp } s$.[28] Because rationalizability of s is equivalent to $\partial\kappa(s)$ being nonempty, Caplin, Dean, and Leahy's (2021) result delivers the following conclusion: whenever Ω is finite and C is posterior separable, $\partial\kappa(s)$ is nonempty if and only if $\partial c(\mu_a^s)$ is nonempty for all $a \in \text{supp } s$.

One can decompose Caplin, Dean, and Leahy's (2021) argument into two. First, they show one can construct a utility function that rationalizes s by setting u_a to be an appropriately normalized member of $\partial c(\mu_a^s)$. In the appendix, we show Caplin, Dean, and Leahy's (2021) construction extends to infinite states and the case in which costs are merely differentiable. In other words, we show $\partial\kappa(s)$ is nonempty whenever C admits a derivative c at p^s for which $\partial c(\mu_a^s)$ is nonempty for all $a \in \text{supp}(s)$.

Caplin, Dean, and Leahy (2021) also establish a converse: when costs are posterior separable and the state is finite, s is rationalizable *only if* $\partial c(\mu_a^s)$ is nonempty for all a in s's support. To prove this claim, Caplin, Dean, and Leahy (2021) prove a duality result to obtain a Kuhn-Tucker-like necessary condition for s to be u-rationalizable, and show adding the relevant multiplier to u_a witnesses $\partial c(\mu_a^s)$ being nonempty. Our results imply this approach generalizes to differentiable costs as well. The reason is that, under Theorem 2's conditions, s is rationalizable if and only if it is rationalizable by C's posterior separable approximation at p^s. Thus, given s and a finite-valued cost that admits a derivative c at p^s, the SCR s is rationalizable only if $\partial c(\mu_a^s)$ is nonempty at all $a \in \text{supp}(s)$—*provided* the state is finite. Finite states are necessary because Caplin, Dean, and Leahy's (2021) duality result may not apply with infinite states. To establish similar duality results for the infinite state case, one usually needs additional regularity conditions (see, e.g., Gretsky, Ostroy, and Zame, 2002; Dworczak and Kolotilin, 2019). Lacking such a result or an alternative proof method, we do not know whether $\partial\kappa(s)$ being nonempty implies the nonemptiness of $\partial c(\mu_a^s)$ for all $a \in \text{supp}(s)$ when the state is infinite.

The construction from the first part of Caplin, Dean, and Leahy's (2021) argument delivers an alternative way of proving the finite-state portion of Proposition 2 for the posterior separable case. Given our above-mentioned result, the same argument extends to costs that are differentiable. For an explanation, recall the subdifferential of a convex function is nonempty over the relative interior of its domain (see, e.g., Rockafellar, 1970, Theorem 23.4), which is always nonempty in finite dimensions. Because an interior s only reveals

[28] See the definition of $\partial c(\mu_a^s)$ in Appendix A.6.

posteriors that have full support, and because all such posteriors are interior when the state is finite, one gets that, when Ω is finite, $\partial c(\mu_a^s) \neq \varnothing$ for all a whenever s is interior. Therefore, when costs are differentiable, one can use Caplin, Dean, and Leahy's (2021) construction to prove Proposition 2 for the finite-state case. However, with infinite states, Caplin, Dean, and Leahy's (2021) approach does not deliver an immediate proof for Proposition 2. The reason is that the relative interior of c's domain is empty, and so $\partial c(\mu_a^s)$ may be empty as well. By focusing on the subdifferential of κ (which is well defined even when C is not differentiable), our proof not only avoids this issue, but also establishes Proposition 2 for a more general class of cost functions.

Convexity and Monotonicity. In our analysis, we assumed C is monotone and convex. We now explain these two assumptions are essentially without loss. In particular, we argue the indirect cost function generated by every lower semicontinuous and proper $\hat{C}$ is identical to the indirect cost generated by a convex and monotone cost function.

To get the result, we must first redefine the agent's indirect cost function so as to allow randomization over information policies. Specifically, we let the agent choose a distribution over information policies, $Q \in \Delta\mathcal{P}$. We say such a distribution **can induce** an SCR s if some action strategy $\alpha : \Delta\Omega \to \Delta A$ is such that, for every $a \in A$ and every event $\tilde{\Omega}$,

$$\int \alpha(a|\mu)\mu(\hat{\Omega})\, p(\mathrm{d}\mu)\, Q(\mathrm{d}p) = \mathbb{E}\left[\mathbf{1}_{\tilde{\Omega}} s_a\right].$$

The indirect cost function is then given by

$$\kappa(s) = \inf_{Q \in \Delta\mathcal{P}} \int \hat{C}(p)\, Q(\mathrm{d}p) \text{ s.t. } Q \text{ can induce } s. \tag{4}$$

Note randomization is not necessary when C is convex, in which case the above reduces to our previous definition of κ. The next result shows a sense in which such convexity always holds.

Proposition 8. *Let κ be the indirect cost function induced by a lower semicontinuous and proper $\hat{C} : \mathcal{P} \to \overline{\mathbb{R}}$. Then, the infimum in (4) is attained. Moreover, κ is also induced by some cost function C that is lower semicontinuous, proper, convex, and monotone.*

Our argument begins by observing Q can induce s if and only if its mean, $q = \int p\, Q(\mathrm{d}p)$, is more informative than s's revealed information policy p^s. It follows the agent's indirect cost function remains unchanged if we replace $\hat{C}$ with a cost function C that assign each p

with the expected cost (under $\hat{C}$) of the cheapest randomization whose mean is more informative than p. We then show Berge's theorem guarantees C is lower semicontinuous and that monotonicity and convexity of C follow from $\succeq$ being a transitive order that respects convex combinations.

Caplin and Dean (2015) use a similar construction to show every behavior generated from costly flexible learning in a finite collection of menus can be rationalized using a convex and monotone cost function. de Oliveira et al. (2017) prove a representation result for preferences over menus with similar implications. In particular, they show one can always take the cost of p to equal the maximum difference between the agent's benefit from using p in some menu and her certainty equivalent for that menu. Moreover, the resulting cost function is the unique cost function that is simultaneously convex, monotone, zero at no information, and consistent with the agent's preferences over menus.

Continuous Choice with Bounded Utilities. Proposition 4 shows the only way to guarantee continuous choice across all objective functions is to require all discontinuous SCRs to have infinite cost. We now explain one can avoid the use of infinite costs if one is willing to require the agent's choice to be continuous for all *bounded* utility functions. In particular, C generates continuous choice for all bounded utility functions if and only if it satisfies an infinite-slope condition.

Proposition 9. *An SCR s is not rationalizable by any bounded utility function if and only if*

$$\inf_{t \in S^\kappa \setminus \{s\}} \frac{\kappa(s) - \kappa(t)}{\|s - t\|_1} = -\infty. \tag{5}$$

In particular, only continuous SCRs are rationalizable by a bounded utility function if and only if (5) *holds for all discontinuous s.*

The result is an immediate consequence of Gale (1967), who shows bounded steepness is a necessary and sufficient condition for the subdifferential of a convex function to contain some linear function that is continuous with respect to a given norm.[29]

Proposition 9's infinite-slope condition is reminiscent of a different condition by Morris and Yang (2021), who introduce the notion of continuous choice to study equilibrium selection in global games. Holding other players' strategies fixed, one can view the problem of each agent in their game as an instance of our model in which $\Omega \subseteq \mathbb{R}$ is an inter-

[29] Using identical reasoning, one can replace $\|\cdot\|_1$ in Proposition 9's statement to analogously characterize which SCRs are rationalizable by other subspaces of utilities. See Online Appendix B.4 for details.

val, $A = \{0, 1\}$, payoffs are bounded, and S^κ is contained in the space of all SCRs for which s_1 is nondecreasing. Morris and Yang (2021) show multiplying κ by a vanishing constant leads to a sharp equilibrium-selection result, provided only SCRs in $S_{\text{AC}} := \{s \in S : \ s_1 \text{ absolutely continuous}\}$ can be rationalized by some monotone and bounded utility function. They also prove this latter property holds whenever κ satisfies a condition called *expensive perfect discrimination*, which states that for every $s \in S^\kappa \setminus S_{\text{AC}}$, the cost function κ exhibits unbounded $\|\cdot\|_1$-steepness in the direction of SCRs in S_{AC}. By contrast, Proposition 9's condition allows κ to exhibit unbounded steepness from *any* direction, and shows allowing for these additional directions results in a condition that is both necessary and sufficient for ruling out discontinuous SCRs as rationalizable by any bounded utility, including utilities that are not monotone.

Costly Stochastic Choice. In our model, we assumed the agent faces a cost to acquire information, which we then used to derive an indirect cost function over the set of SCRs. By contrast, some models formulate a cost function $\tilde{\kappa}$ on SCRs directly, without micro-founding it via information acquisition (e.g., Mattsson and Weibull, 2002; Fosgerau et al., 2020; Flynn and Sastry, 2021; Morris and Yang, 2021). Some of our results apply to those models as well, provided $\tilde{\kappa}$ is convex, proper, and weak* lower semicontinuous. Because the arguments for Propositions 1 and 2 rely only on properties of the agent's indirect cost function, both of these propositions also apply to models in which the cost of an SCR does not originate from information acquisition. The same holds for parts (i) and (ii) of Proposition 4, as well as Proposition 9.

To get an analogue of Theorem 1, the cost $\tilde{\kappa}$ must satisfy additional properties. The most obvious such property is strict convexity: if $\tilde{\kappa}$ is strictly convex, every rationalizable SCR is uniquely rationalizable, and so the logic behind Proposition 2 delivers a uniformly dense set of SCRs that are uniquely rationalizable.

With strict convexity, one can also get an analogue of Theorem 3, provided $\tilde{\kappa}$ is sufficiently smooth. Say $u \in \mathcal{U}$ is a derivative of $\tilde{\kappa}$ at s if for all t,

$$d_s^+ \tilde{\kappa}(t) = \mathbb{E}\left[u \cdot (t - s)\right].$$

Similar to the case in which C is iteratively differentiable, we show in the appendix that given an interior s, the cost $\tilde{\kappa}$ admits u as a derivative at s only if all utilities that rationalize s differ from u by a nuisance term; that is, v rationalizes s if and only if some $\lambda \in L^1(\mu_0)$ is such that $v_a = u_a + \lambda$ for all a. Armed with this observation, one can repeat the arguments guaranteeing Theorem 3 to show a weak*-dense set of SCRs exists that induce unique subset predictions, provided $\tilde{\kappa}$ is finite-valued, admits a derivative at any interior SCR, and has infinite slope at the edges. Thus, whereas $\tilde{\kappa}$ imposes few restrictions on the agent's behavior in a given menu, across menus, one can still use $\tilde{\kappa}$ to make meaningful predictions about the agent's choices.

References

Aliprantis, Charalambos D and Kim Border. 2006. *Infinite Dimensional Analysis: A Hitchhiker's Guide*. Springer Science & Business Media.

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

Benoît, Jean-Pierre and Juan Dubra. 2011. "Apparent Overconfidence." *Econometrica* 79 (5):1591–1625.

Benyamini, Yoav and Joram Lindenstrauss. 1998. *Geometric Nonlinear Functional Analysis*, vol. 48. American Mathematical Soc.

Blackwell, David. 1953. "Equivalent Comparisons of Experiments." *The Annals of Mathematical Statistics* 24 (2):265–272.

Bloedel, Alexander W and Weijie Zhong. 2020. "The Cost of Optimally-Acquired Information." *Unpublished Manuscript, November* .

Borwein, Jonathan M and Jon D Vanderwerff. 2010. *Convex Functions: Constructions, Characterizations and Counterexamples*, vol. 172. Cambridge University Press Cambridge.

Brøndsted, Arne and Ralph Tyrrell Rockafellar. 1965. "On the Subdifferentiability of Convex Functions." *Proceedings of the American Mathematical Society* 16 (4):605–611.

Caplin, Andrew, Dániel Csaba, John Leahy, and Oded Nov. 2020. "Rational Inattention, Competitive Supply, and Psychometrics." *The Quarterly Journal of Economics* 135 (3):1681–1724.

Caplin, Andrew and Mark Dean. 2013. "Behavioral Implications of Rational Inattention with Shannon Entropy." Tech. rep., National Bureau of Economic Research.

———. 2015. “Revealed Preference, Rational Inattention, and Costly information acquisition.” *American Economic Review* 105 (7):2183–2203.

Caplin, Andrew, Mark Dean, and John Leahy. 2017. “Rationally Inattentive Behavior: Characterizing and Generalizing Shannon Entropy.” Tech. rep., National Bureau of Economic Research.

———. 2019. “Rational Inattention, Optimal Consideration Sets, and Stochastic Choice.” *The Review of Economic Studies* 86 (3):1061–1094.

———. 2021. “Rationally Inattentive Behavior: Characterizing and Generalizing Shannon Entropy.” Tech. rep., National Bureau of Economic Research.

Cerreia-Vioglio, Simone, Fabio Maccheroni, and Massimo Marinacci. 2017. “Stochastic Dominance Analysis Without the Independence Axiom.” *Management Science* 63 (4):1097–1109.

Chambers, Christopher P, Ce Liu, and John Rehbeck. 2020. “Costly Information Acquisition.” *Journal of Economic Theory* 186:104979.

Csiszár, Imre. 1974. “On an Extremum Problem of Information Theory.” *Studia Scientiarum Mathematicarum Hungarica* 9:57–71.

de Oliveira, Henrique. 2014. “Axiomatic Foundations for Entropic Costs of Attention.” Tech. rep., 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.

Dean, Mark and Nathaniel Neligh. 2019. “Experimental Tests of Rational Inattention.” .

Denti, Tommaso. 2022. “Posterior-Separable Cost of Information.” Tech. rep., working paper.

Denti, Tommaso, Massimo Marinacci, and Luigi Montrucchio. 2020. “A Note on Rational Inattention and Rate Distortion Theory.” *Decisions in Economics and Finance* 43 (1):75–89.

Denti, Tommaso, Massimo Marinacci, Aldo Rustichini et al. 2021. *Experimental cost of information*. IGIER, Università Bocconi.

Dewan, Ambuj and Nathaniel Neligh. 2020. "Estimating Information Cost Functions in Models of Rational Inattention." *Journal of Economic Theory* 187:105011.

Dillenberger, David, Vijay Krishna, and Philipp Sadowski. 2020. "Subjective Information Choice Processes." .

Dworczak, Piotr and Anton Kolotilin. 2019. "The Persuasion Duality." *Available at SSRN 3474376* .

Flynn, Joel P and Karthik Sastry. 2021. "Strategic Mistakes." *Available at SSRN 3663481* .

Fosgerau, Mogens, Emerson Melo, Andre De Palma, and Matthew Shum. 2020. "Discrete Choice and Rational Inattention: A General Equivalence Result." *International economic review* 61 (4):1569–1589.

Frankel, Alexander and Emir Kamenica. 2019. "Quantifying Information and Uncertainty." *American Economic Review* 109 (10):3650–80.

Gale, David. 1967. "A Geometric Duality Theorem with Economic Applications." *The Review of Economic Studies* 34 (1):19–24.

Gretsky, Neil E, Joseph M Ostroy, and William R Zame. 2002. "Subdifferentiability and The Duality Gap." *Positivity* 6 (3):261–274.

Hébert, Benjamin and Michael Woodford. 2021a. "Neighborhood-Based Information Costs." *American Economic Review* 111 (10):3225–55.

———. 2021b. "Rational Inattention when Decisions Take Time." .

Hong, Chew Soo, Edi Karni, and Zvi Safra. 1987. "Risk Aversion in the Theory of Expected Utility with Rank Dependent Probabilities." *Journal of Economic theory* 42 (2):370–381.

Hunt, Brian R, Tim Sauer, and James A Yorke. 1992. "Prevalence: A Translation-Invariant "Almost Every" on Infinite-Dimensional Spaces." *Bulletin of the American mathematical society* 27 (2):217–238.

Kallenberg, Olav. 2017. *Random Measures, Theory and Applications*, vol. 1. Springer.

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

Khan, Ali, Haomiao Yu, Zhixiang Zhang et al. 2019. “Information Structures on a General State Space: An Equivalence Theorem and an Application.” Tech. rep.

Lin, Yi-Hsuan. 2022. “Stochastic Choice and Rational Inattention.” *Journal of Economic Theory* :105450.

Lipnowski, Elliot and Laurent Mathevet. 2018. “Disclosure to a Psychological Audience.” *American Economic Journal: Microeconomics* 10 (4):67–93.

Matějka, Filip and Alisdair McKay. 2015. “Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model.” *American Economic Review* 105 (1):272–98.

Mattsson, Lars-Göran and Jörgen W Weibull. 2002. “Probabilistic Choice and Procedurally Bounded Rationality.” *Games and Economic Behavior* 41 (1):61–78.

Mensch, Jeffrey. 2018. “Cardinal Representations of Information.” *Available at SSRN 3148954* .

Morris, Stephen and Philipp Strack. 2019. “The Wald Problem and the Relation of Sequential Sampling and Ex-Ante Information Costs.” *Available at SSRN 2991567* .

Morris, Stephen and Ming Yang. 2021. “Coordination and Continuous Stochastic Choice.” .

Phelps, Robert R. 2001. *Lectures on Choquet’s theorem*. Springer Science & Business Media.

———. 2009. *Convex Functions, Monotone Operators and Differentiability*, vol. 1364. Springer.

Pomatto, Luciano, Philipp Strack, and Omer Tamuz. 2020. “The Cost of Information.” .

Posner, Edward. 1975. “Random Coding Strategies for Minimum Entropy.” *IEEE Transactions on Information Theory* 21 (4):388–391.

Pourciau, BH. 1983. “Multiplier Rules and the Separation of Convex Sets.” *Journal of Optimization Theory and Applications* 40 (3):321–331.

Rockafellar, Ralph Tyrell. 1970. *Convex Analysis*. Princeton University Press.

Sims, Christopher A. 1998. “Stickiness.” In *Carnegie-Rochester Conference Series on Public Policy*, vol. 49. Elsevier, 317–356.

———. 2003. "Implications of Rational Inattention." *Journal of Monetary Economics* 50 (3):665–690.

———. 2006. "Rational Inattention: Beyond the Linear-Quadratic Case." *American Economic Review* 96 (2):158–163.

Srivastava, Sashi Mohan. 2008. *A Course on Borel Sets*, vol. 180. Springer Science & Business Media.

Winkler, Gerhard. 1988. "Extreme Points of Moment Sets." *Mathematics of Operations Research* 13 (4):581–587.

A. Proof Appendix

A.1. Section 3 Proofs

Proof of Lemma 1. First, suppose $p \succeq p^s$, as witnessed by $m : \Delta\Omega \to \Delta\Delta\Omega$. Any $a \in \text{supp}(s)$ then has $m(\cdot|\mu_a^s) \ll p$ because p is a proper weighted average of the finitely many measures $\{m(\cdot|\mu_a^s)\}_{a\in\text{supp}(s)}$. So let $\alpha_a : \Delta\Omega \to [0,1]$ be some version of the scaled Radon-Nikodym derivative $p_a^s \, \frac{\mathrm{d}m(\cdot|\mu_a^s)}{\mathrm{d}p}$ for each $a \in A$ with $p_a^s > 0$; and let $\alpha_a = \mathbf{0}$ for every other $a \in A$. By construction, $\sum_{a\in A} \alpha_a =_{p\text{-a.e.}} \mathbf{1}$, so we can change $\{\alpha_a\}_{a\in A}$ on a p-null set to ensure the equation holds globally. Let us now show the strategy (p,α) induces s, where $\alpha := \sum_{a\in A} \alpha_a \delta_a$. Indeed, for any action $a \in A$ and event $\hat{\Omega} \subseteq \Omega$, the strategy's induced probability of action a being played and event $\hat{\Omega}$ occurring is zero (like under s) if p_a^s is zero, and is otherwise equal to

$$\begin{aligned}\int \mu(\hat{\Omega})\alpha_a(\mu)\, p(\mathrm{d}\mu) &= \int \mu(\hat{\Omega}) p_a^s \, \tfrac{\mathrm{d}r(\cdot|\mu_a^s)}{\mathrm{d}p}(\mu)\, p(\mathrm{d}\mu) \\ &= p_a^s \int \mu(\hat{\Omega})\, m(\mathrm{d}\mu|\mu_a^s) \\ &= p_a^s \mu_a^s(\hat{\Omega}) \\ &= \mathbb{E}\left[\mathbf{1}_{\hat{\Omega}}\, s_a\right].\end{aligned}$$

Therefore, p can induce s.

Conversely, suppose some strategy (p,α) induces s. For any $a \in A$ with $p_a^s > 0$, we can define $q^a \in \Delta\Delta\Omega$ by letting $q^a(D) := \frac{1}{p_a^s} \int_D \alpha(a|\mu)\, p(\mathrm{d}\mu)$ for every Borel $D \subseteq \Delta\Omega$. Every

$a \in \text{supp}(s)$ and every event $\hat{\Omega} \subseteq \Omega$ then have

$$\int \mu(\hat{\Omega})\, q^a(\mathrm{d}\mu) = \tfrac{1}{p_a^s} \int \mu(\hat{\Omega})\alpha(a|\mu)\, p(\mathrm{d}\mu) = \tfrac{1}{p_a^s} \int_{\hat{\Omega}} s_a(\omega)\, \mu_0(\mathrm{d}\omega) = \mu_a^s(\hat{\Omega}).$$

Said differently, every $a \in \text{supp}(s)$ has $\int \mu\, q^a(\mathrm{d}\mu) = \mu_a^s$. Hence, $p = \sum_{a\in A} p_a^s q^a \succeq \sum_{a\in A} p_a^s \delta_{\mu_a^s} = p^s$. □

Proof of Lemma 2, part (i). Recall $s^n \to s$ weak* in S tells us

$$p_a^{s^n} \int f(\omega)\, \mu_a^{s^n}(\mathrm{d}\omega) \to p_a^s \int f(\omega)\, \mu_a^s(\mathrm{d}\omega) \text{ as } n \to \infty,\ \forall a \in A \text{ and } f \in L^1(\mu_0).$$

Consequently, any $a \in A$ (specializing to $f = \mathbf{1}$) has $p_a^{s^n} \to p_a^s$; and any $a \in \text{supp}(s)$—scaling a given $f \in L^1(\mu_0)$ by $\frac{p_a^{s^n}}{p_a^s}$, which converges to 1—has $\int f(\omega)\, \mu_a^{s^n}(\mathrm{d}\omega) \to \int f(\omega)\, \mu_a^s(\mathrm{d}\omega)$ for every $f \in L^1(\mu_0)$. Because every continuous $: \Omega \to \mathbb{R}$ represents some element of $L^1(\mu_0)$, the latter property tells us $\mu_a^{s^n} \to \mu_a^s$ in $\Delta\Omega$ if $a \in \text{supp}(s)$. Hence, $p^{s^n} \to p^s$ in $\mathcal{P}$, as desired. □

For the proof that follows, recall the Hardy-Littlewood-Polya-Blackwell-Stein-Sherman-Cartier theorem (Phelps, 2001, p. 94)—hereafter, the **HLPBSSC theorem**—says $p, q \in \mathcal{P}$ satisfy $p \succeq q$ if and only if $\int f\, p(\mathrm{d}\mu) \geq \int f\, q(\mathrm{d}\mu)$ for every convex continuous $f : \Delta\Omega \to \mathbb{R}$.

Proof of Lemma 2, parts (ii) and (iii). Let $r := (1-\beta)s + \beta t$ and $p := (1-\beta)p^s + \beta p^t$ Direct computation shows $\text{supp}(r) = \text{supp}(s) \cup \text{supp}(t)$ and, for every $a \in \text{supp}(r)$,

$$\begin{aligned} p_a^r &= (1-\beta)p_a^s + \beta p_a^t \\ \mu_a^r &= \tfrac{(1-\beta)p_a^s}{p_a^r}\mu_a^s + \tfrac{\beta p_a^t}{p_a^r}\mu_a^t. \end{aligned}$$

Observe now that any convex continuous $f : \Delta\Omega \to \mathbb{R}$ has

$$\begin{aligned} \int f(\omega)\, p(\mathrm{d}\omega) - \int f(\omega)\, p^r(\mathrm{d}\omega) &= \sum_{a\in\text{supp}(r)} \left[(1-\beta)p_a^s f(\mu_a^s) + \beta p_a^t f\left(\mu_a^t\right) - p_a^r f(\mu_a^r)\right] \\ &= \sum_{a\in\text{supp}(r)} p_a^r \left[\tfrac{(1-\beta)p_a^s}{p_a^r} f(\mu_a^s) + \tfrac{\beta p_a^t}{p_a^r} f\left(\mu_a^t\right) - f(\mu_a^r)\right] \\ &\geq \sum_{a\in\text{supp}(r)} p_a^r \left[f\left(\tfrac{(1-\beta)p_a^s}{p_a^r}\mu_a^s + \tfrac{\beta p_a^t}{p_a^r}\mu_a^t\right) - f(\mu_a^r)\right] \\ &= \sum_{a\in\text{supp}(r)} p_a^r \left[f(\mu_a^r) - f(\mu_a^r)\right] \\ &= 0. \end{aligned}$$

The HLPBSSC theorem then implies $p \succeq p^r$, delivering part (ii).

Toward (iii), suppose some $a \in \text{supp}(s) \cap \text{supp}(t)$ is such that $\mu_a^s \neq \mu_a^t$. Now, specialize the above algebra to the case in which $f|_{\text{co}\{\mu_a^s, \mu_a^t\}}$ is strictly convex.[30] Then, the inequality in the above chain is strict, witnessing $\int f(\omega)\, p(\mathrm{d}\omega) - \int f(\omega)\, p^r(\mathrm{d}\omega) > 0$ so that $p \succ p^r$. The result follows. □

A.2. On the Value Function

Extend κ to $\mathcal{S}$ by setting $\kappa(s) = \infty$ for all $s \in \mathcal{S} \setminus S$. The goal of this section is to prove some results regarding the optimal **value function** $V : \mathcal{U} \to \mathbb{R}$ defined as

$$V(u) := \max_{s \in \mathcal{S}} \left[\mathbb{E}[u \cdot s] - \kappa(s)\right] = \max_{s \in S} \left[\mathbb{E}[u \cdot s] - \kappa(s)\right].$$

Our results also pertain to the subdifferential of V at a utility u,

$$\begin{aligned} \partial V : \mathcal{U} &\rightrightarrows S, \\ u &\mapsto \{s \in S :\ V(v) \geq V(u) + \mathbb{E}\left[(v-u)s\right] \text{ for all } v \in \mathcal{U}\}. \end{aligned}$$

For any subset $T \subseteq S$, define the upper and lower inverses of ∂V:

$$\begin{aligned} \partial V^U(T) &:= \{u \in \mathcal{U} :\ \partial V(u) \subseteq T\}, \\ \partial V^L(T) &:= \{u \in \mathcal{U} :\ \partial V(u) \cap T \neq \varnothing\}. \end{aligned}$$

We say ∂V is **norm-to-norm (resp. norm-to-weak*) upper hemicontinuous** if $\partial V^U(T)$ is norm open whenever T is norm (resp. weak*) open.

The following lemma establishes useful continuity properties of the optimal value function and its subdifferential.

Lemma 3. *The value function $V : \mathcal{U} \to \mathbb{R}$ is convex and norm continuous, and its subdifferential is norm-to-weak* upper hemicontinuous.*

Proof. The value function is real-valued because (as noted in the main text as a consequence of Banach-Alaoglu) each $u \in \mathcal{U}$ admits some u-rationalizable SCR. It is convex as a supremum of affine functions.

[30] For instance, f could be given by $f(\mu) := \left[\int g(\omega)\, \mu(\mathrm{d}\omega)\right]^2$ for some continuous $g : \Omega \to \mathbb{R}$ with $\int g(\omega)\, \mu_a^s(\mathrm{d}\omega) \neq \int g(\omega)\, \mu_a^t(\mathrm{d}\omega)$.

To show V is norm-continuous, we need only show (Aliprantis and Border, 2006, Theorem 5.43) it is bounded above on some ball. And indeed, any $u \in \mathcal{U}$ with $\|u\|_1 \leq 1$ has

$$V(u) \leq \max_{s \in \mathcal{S}\colon \kappa(s)<\infty} \mathbb{E}[u \cdot s] \leq \|u\|_1 \max_{s \in S} \|s\|_\infty = \|u\|_1 \leq 1.$$

Finally, upper hemicontinuity of ∂V then follows from Proposition 6.1.1 of Borwein and Vanderwerff (2010). □

Next, we collect some standard facts from convex analysis, applied directly to our setting. To state them, given $s \in S^\kappa$ and $s' \in S$, let $d_s^+\kappa$ denote the **directional derivative** of κ at s in direction $s' - s$.

Lemma 4. *Viewing $\mathcal{U}$ with its norm topology and $\mathcal{S}$ with its weak* topology, V is the convex conjugate of κ, and κ is the convex conjugate of V. Moreover, the following are equivalent:*

(i) $s \in \operatorname{argmax}_{t \in S} \left[\mathbb{E}[u \cdot t] - \kappa(t)\right]$.

(ii) $u \in \partial\kappa(s)$.

(iii) Every $s' \in S^\kappa$ *has* $d_s^+\kappa(s') \geq \mathbb{E}\left[u \cdot (s' - s)\right]$.

(iv) $s \in \partial V(u)$.

Proof. Recall $\mathcal{U}$ with its norm topology and $\mathcal{S}$ with its weak* topology form a dual pair with the bilinear map $(u,t) \mapsto \mathbb{E}\left[u \cdot t\right]$. With these respective topologies, κ is proper, convex, and lower semicontinuous (by Corollary 1). By definition of V, it equals the convex conjugate of the indirect cost function; that is, $V = \kappa^*$. Hence, by the Fenchel-Moreau theorem (e.g., Borwein and Vanderwerff, 2010, Proposition 4.4.2), κ is the convex conjugate of V; that is, $\kappa = V^*$.

Now, we pursue the four-way equivalence. First, that (i) is equivalent to (ii) is immediate (see discussion after the statement of Proposition 1). Next, Aliprantis and Border's (2006) Theorem 7.16 directly implies (ii) is equivalent to (iii). Finally, to show (i) is equivalent to (iv), Borwein and Vanderwerff's (2010) Proposition 4.4.1 part (a) delivers that $s \in \partial V(u)$ holds if and only if

$$V(u) + \kappa(s) = \mathbb{E}[u \cdot s],$$

which is equivalent to

$$\mathbb{E}[u \cdot s] - \kappa(s) = V(u) = \max_{t \in S} \left[\mathbb{E}[u \cdot t] - \kappa(t)\right].$$

It follows (iv) is equivalent to (i), as desired. ◻

The following lemma provides sufficient conditions for a set of SCRs to have its rationalizing utilities be an open set.

Lemma 5. *Suppose $T \subseteq S$ is weak* open, and every $u \in \mathcal{U}$ and u-rationalizable $s \in T$ are such that s is uniquely u-rationalizable. Then, the set of utilities that rationalize SCRs in T is open in $\mathcal{U}$.*

Proof. Recall $u \in \mathcal{U}$ rationalizes $s \in S$ if and only if $s \in \partial V(u)$ (Lemma 4). It follows the set of utilities that rationalize the SCRs in T is given by $\partial V^L(T)$. Moreover, because ∂V is norm-to-weak* upper hemicontinuous (Lemma 3) and T is weak* open, we know $\partial V^U(T)$ is norm open. Hence, the lemma will follow if we can establish that $\partial V^L(T) = \partial V^U(T)$.

Toward showing $\partial V^L(T) = \partial V^U(T)$, we use the fact that u rationalizes $s \in T$ if and only if s is uniquely u-rationalizable. Therefore, $\partial V(u) \cap \{s\} \neq \varnothing$ if and only if $\partial V(u) \subseteq \{s\}$, meaning $\partial V^L(\{s\}) = \partial V^U(\{s\})$ must hold for all $s \in T$. As such,

$$\partial V^L(T) = \cup_{s\in T}\partial V^L(\{s\}) = \cup_{s\in T}\partial V^U(\{s\}) \subseteq \partial V^U(T) \subseteq \partial V^L(T),$$

where the last containment follows from ∂V being nonempty-valued. It follows $\partial V^L(T) = \partial V^U(T)$. ◻

A.3. Section 4 Proofs

Proof of Proposition 1. Let $\mathcal{V}$ denote the set of $u \in \mathcal{U}$ at which V is Gâteaux differentiable, and recall V is continuous by Lemma 3. For any $u \in \mathcal{V}$, the function V has a unique subgradient at u (Borwein and Vanderwerff, 2010, Corollary 4.2.5), and so a unique SCR is u-rationalizable.

It therefore remains to show $\mathcal{U} \setminus \mathcal{V}$ is meager and shy. That this set is meager follows from Mazur's theorem (see Borwein and Vanderwerff, 2010, Theorem 4.6.3), which tells us $\mathcal{V}$ is dense and G_δ, and hence co-meager. To see $\mathcal{U} \setminus \mathcal{V}$ is shy, it suffices to show it is Haar null.[31]

To show $\mathcal{U}\setminus\mathcal{V}$ is Haar null, note Theorem 5.44 from Aliprantis and Border (2006) tells us V is locally Lipschitz. Hence, being second countable, $\mathcal{U}$ can be covered by countably many

[31] Given Fact 2 from Hunt, Sauer, and Yorke (1992) (resp. Proposition 4.6.1(c) from Borwein and Vanderwerff, 2010), $\mathcal{U} \setminus \mathcal{V}$ is shy (resp. Haar null) if and only if some compactly supported finite Borel (resp. Radon) measure assigns zero measure to every translation of it. Because every finite Borel measure on the Polish space $\mathcal{U}$ is a Radon measure, $\mathcal{U} \setminus \mathcal{V}$ is shy if and only if it is Haar null.

open balls $\{\mathcal{U}_n\}_{n=1}^{\infty}$ with $V|_{\mathcal{U}_n}$ Lipschitz for each $n \in \mathbb{N}$. Moreover, for each $n \in \mathbb{N}$, Theorem 4.6.5 from Borwein and Vanderwerff (2010) tells us $\mathcal{U}_n \setminus \mathcal{V}$ is Aronszajn null, and hence (by Benyamini and Lindenstrauss, 1998, Proposition 6.25) Haar null. Therefore, $\mathcal{U} \setminus \mathcal{V} = \bigcup_{n=1}^{\infty}(\mathcal{U}_n \setminus \mathcal{V})$ is Haar null (by Borwein and Vanderwerff, 2010, Proposition 4.6.1(e)). □

A.4. Section 5 Proofs

Proof of Proposition 2. Recall from Lemma 4 that $s \in S$ is rationalizable if and only if κ is subdifferentiable at s, that is, if and only if $\partial\kappa(s) \neq \varnothing$. Also from that lemma, κ is the convex conjugate of the function V defined on the Banach space $\mathcal{U}$. Moreover, V is proper, convex, and continuous (Lemma 3). Therefore, the dual version of the Brøndsted-Rockafellar theorem (Brøndsted and Rockafellar, 1965, second part of Theorem 2) says κ is subdifferentiable on a norm-dense subset of its domain S^{κ}, as required.

The second point follows from the fact (Theorem 23.4 in Rockafellar, 1970) that a proper convex function on a Euclidean space is subdifferentiable everywhere in the relative interior of its domain. □

The following lemma shows κ is "almost" strictly convex if no marginal information can be acquired for free.

Lemma 6. *If C is strictly monotone, and $s, t \in S$ and $a \in A$ are such that s_a and t_a are not proportional, then κ is strictly convex on* $\mathrm{co}\{s, t\}$.

Proof. Recall Lemma 1 tells us $\kappa(r) = C(p^r)$ for every $r \in S$.

Take any $\beta \in (0, 1)$, and let $r := (1-\beta)s + \beta t$ and $p := (1-\beta)p^s + \beta p^t$. Part (iii) from Lemma 2 shows $p \succ p^r$. Hence,

$$\kappa(r) = C(p^r) < C(p) \leq (1-\beta)C(p^s) + \beta C(p^t) = (1-\beta)\kappa(s) + \beta\kappa(t),$$

where the inequalities come from strict monotonicity and convexity of C, respectively. □

Proof of Proposition 3. The second assertion follows immediately from the first, because the expected benefit from a stochastic choice rule is an affine function of the stochastic choice rule, and a strictly concave function can have at most one maximizer. We turn now to the first assertion.

Given that κ is already known to be weakly convex by Corollary 1, we need only show, given arbitrary $t \in S \setminus \{s\}$, that κ is strictly convex on $\mathrm{co}\{s, t\}$.

So suppose $t \in S$ is such that κ is not strictly convex on $\mathrm{co}\{s,t\}$. Lemma 6 then tells us t_a is a scalar multiple of s_a (which is assumed to be nonzero) for every $a \in A$. Equivalently, $\mu_a^t = \mu_a^s$ for every $a \in \mathrm{supp}(t)$. Hence,

$$\sum_{a\in A} p_a^t \mu_a^s = \sum_{a\in A} p_a^t \mu_a^t = \mu_0 = \sum_{a\in A} p_a^s \mu_a^s.$$

Affine independence of the $|A|$ beliefs $\{\mu_a^s\}_{a\in A}$ then implies $p_a^t = p_a^s$ for every $a \in A$, so that $t = s$, delivering the result. □

Although not relevant to our subsequent results, we briefly note a stronger uniqueness property (proven in Appendix C) that follows readily for the special case of binary actions.

Corollary 2. *Suppose $|A| = 2$ and C is strictly monotone, and fix $u \in \mathcal{U}$. Either a unique SCR is u-rationalizable, or every u-rationalizable SCR generates state-independent behavior. In particular, all optimal strategies entail the same information policy.*

The following lemma formalizes a sense in which the affine independence case of Proposition 3 is the typical case.

Lemma 7. *If $|\Omega| \geq |A|$, then $\mathrm{supp}(p^s)$ consists of $|A|$ affinely independent beliefs—that is, s is linearly independent—for a weak*-open and uniformly dense set of $s \in S$.*

Proof. Define $\hat{\mathcal{S}}$ to be the set of all $s \in \mathcal{S}$ such that $\sum_{a\in A} s_a = \mathbf{1}$.

Let $\Omega = \bigsqcup_{a\in A} \Omega_a$ be a Borel partition into sets with positive μ_0-measure; such a partition exists because Ω is metrizable with at least $|A|$ distinct points and μ_0 has full support. Then, define the set $\hat{M} \subseteq \mathbb{R}^{A\times A}$ as the set of matrices for which the a row's entries sum to $\mu_0(\Omega_a)$ for each $a \in A$, and the function $\pi : \hat{\mathcal{S}} \to \hat{M}$ given by $\pi(s) := \left[\int_{\Omega_a} s_{\tilde{a}}(\omega)\, \mu_0(\mathrm{d}\omega)\right]_{a,\tilde{a}\in A}$. The map π is affine, weak* continuous, and hence norm continuous, and, since $\{\Omega_a\}_{a\in A}$ are pairwise disjoint with nonzero measure, surjective. Because both $\hat{\mathcal{S}}$ and $\hat{M}$ are closed affine subspaces of Banach spaces, it follows from the open mapping theorem that π is a norm-open map: it maps norm-open sets to open sets. Hence, if G is any open subset of $\hat{M}$ with closure equal to $\pi(S)$, then $\pi^{-1}(G)$ is weak* open in $\hat{\mathcal{S}}$ (because π is weak* continuous) with norm closure containing S (because π is norm open).[32] It therefore suffices to find some

[32] To see the latter implication, let T be any norm-open set in $\hat{\mathcal{S}}$ that intersects S. Observe that because T intersects S, the set $\pi(T)$ intersects $\pi(S)$ too. Note also that $\pi(T)$ is open in $\hat{M}$, because π is norm open. Therefore, $\pi(S)$ is contained in the norm closure of G, and so G intersects $\pi(T)$. Said differently, T intersects $\pi^{-1}(G)$, as required.

open $G \subseteq \hat{M}$ with closure equal to $\pi(S)$ such that any given $s \in S \cap \pi^{-1}(G)$ has $\operatorname{supp}(p^s)$ consisting of $|A|$ affinely independent beliefs.

To that end, let G be the convex set of invertible matrices in $\hat{M}$ with strictly positive entries. The set G is open in $\hat{M}$ because the determinant function is continuous. Moreover, it is contained in $\pi(S)$, which is the set of matrices in $\hat{M}$ with nonnegative entries. Next, to see G is nonempty, define $g_\epsilon := \left[\epsilon + (\mu_0(\Omega_a) - |A|\epsilon)\, \mathbf{1}_{a=\tilde{a}}\right]_{a,\tilde{a}\in A} \in \hat{M}$ for $\epsilon \in \left[0, \frac{\min_{a\in A}\mu_0(\Omega_a)}{|A|}\right]$. Because g_0 is diagonal with nonzero diagonal entries, its determinant is nonzero. Thus, since a nonzero univariate polynomial has only finitely many roots, $\det(g_\epsilon)$ is nonzero for $\epsilon > 0$ sufficiently small. The matrix $g := g_\epsilon$ belongs to G for such an ϵ. Now, we observe G is dense in $\pi(S)$. Fixing an arbitrary $h \in \pi(S)$, we want to show h is a limit of matrices from G. Define the function $\mathbb{R} \to \mathbb{R}$ via $\gamma \mapsto \det\left[\gamma g + (1-\gamma)h\right]$ and note it is a polynomial that is not globally zero (by evaluating at $\gamma = 1$). Hence, the polynomial has only finitely many roots. In particular, some $\bar{\gamma} \in (0,1)$ exists such that $\gamma g + (1-\gamma)h \in \hat{M}$ is invertible for every $\gamma \in (0, \bar{\gamma})$—and so h is in the closure of $\hat{M}$.

All that remains is to show, for any given $s \in S \cap \pi^{-1}(G)$, that $\operatorname{supp}(p^s)$ consists of $|A|$ affinely independent beliefs. Toward showing this property, observe $p^s_a > 0$ for every $a \in A$ because $\pi(s)$ has nonnegative entries and no zero columns. We can now define the matrix $m := \left[\frac{1}{p^s_{\tilde{a}}}\int_{\Omega_a} s_{\tilde{a}}\,\mathrm{d}\mu_0\right]_{a,\tilde{a}\in A} = \left[\mu^s_{\tilde{a}}(\Omega_a)\right]_{a,\tilde{a}\in A}$. Because m is a product of $\pi(s)$ and another invertible matrix, its columns are linearly independent, and therefore affinely independent. It follows directly that the $|A|$ vectors $\{\mu^s_a\}_{a\in A}$ are affinely independent, as required. $\square$

Proof of Claim 1. Take $u \in \mathcal{U}$ and $s \in S$. Suppose s is the unique stochastic choice rule induced by some optimal strategy given u. We must show $|\operatorname{supp}(s)| \leq |\Omega|$. The result is vacuous (given $|A| < \infty$) when Ω is infinite, so we focus on the case in which Ω is finite.

First, let us establish s_a and $s_{\tilde{a}}$ cannot be proportional for any two distinct $a, \tilde{a} \in \operatorname{supp}(s)$. For a contradiction, assume they are in fact proportional. For any $\beta \in [0,1]$, then, we can define $s^\beta \in S$ via $s^\beta_a := (1-\beta)(s_a + s_{\tilde{a}})$, $s^\beta_{\tilde{a}} := \beta(s_a + s_{\tilde{a}})$, and $s^\beta_{a'} := s_{a'}$ for every $a' \in A\backslash\{a, \tilde{a}\}$. By construction, p^{s^β} is the same for every $\beta \in [0,1]$, so that Lemma 1 implies $\kappa(s^\beta)$ is the same for every $\beta \in [0,1]$. Therefore, the objective $\beta \mapsto \mathbb{E}\left[u \cdot s^\beta\right] - \kappa(s^\beta)$ is affine because $\beta \mapsto s^\beta$ is. It follows this objective cannot be uniquely maximized at $\beta^s = \frac{p^s_{\tilde{a}}}{p^s_a + p^s_{\tilde{a}}} \in (0,1)$—contradicting unique optimality of s because different values of β generate different stochastic choice rules.

Having shown s_a and $s_{\tilde{a}}$ cannot be proportional for any two distinct $a, \tilde{a} \in \operatorname{supp}(s)$, we know $\{\mu^s_a\}_{a\in\operatorname{supp}(s)}$ are $|\operatorname{supp}(s)|$ distinct elements of $\Delta\Omega$. No more than $|\Omega|$ affinely independent vectors can exist in a set (like $\Delta\Omega$) with dimension $|\Omega| - 1$, so the claim will follow if we show $\{\mu_a^s\}_{a\in\text{supp}(s)}$ are affinely independent.

Assume, for a contradiction, $\text{supp}(p^s) = \{\mu_a^s\}_{a\in\text{supp}(s)}$ are affinely dependent. By Winkler's (1988) Theorem 2.1(b), we then have $p^s \notin \text{ext}(\mathcal{P})$. That is, p^s is the midpoint of two distinct information policies $p^1, p^2 \in \mathcal{P}$. Moreover, replacing each of p^1, p^2 with a weighted average with p^s if necessary, we may assume C is affine on $\text{co}\{p^1, p^2\}$. Optimality of s implies the agent's optimal value is

$$\mathbb{E}\left[\sum_{a\in\text{supp}(s)} u_a s_a\right] - C(p^s) = \sum_{i=1,2} \tfrac{1}{2}\left\{\left[\sum_{a\in\text{supp}(s)} p^i(\mu_a^s)\int u_a \mathrm{d}\mu_a^s\right] - C(p^i)\right\}.$$

Hence, some $i \in \{1,2\}$ has $v := \left[\sum_{a\in\text{supp}(s)} p^i(\mu_a^s)\int u_a \mathrm{d}\mu_a^s\right] - C(p^i)$ weakly higher than the agent's optimal value. Payoff v is clearly attainable via some strategy that induces stochastic choice rule $s^i \in S$ given by $s_a^i := \frac{p^i(\mu_a^s)}{p_a^s} s_a$ for $a \in \text{supp}(s)$ and $s_a^i := 0$ for $a \in A \setminus \text{supp}(s)$. Hence, s^i is rationalizable, in contradiction to s being uniquely so. The claim follows. □

Proof of Theorem 1. Let $R \subseteq S$ be the set of all rationalizable SCRs if Ω is infinite, and let it be the relative interior of S^κ (which is the interior of S^κ in S by Assumption A1(iii)) if Ω is finite. By Proposition 2, R is norm dense (i.e., uniformly dense) in S^κ, and every SCR in R is rationalizable. So, R is a norm-dense subset of S^κ consisting only of rationalizable SCRs, and R is open in S if Ω is finite.

Now, Lemma 7 delivers a weak*-open (and hence norm-open) and norm-dense subset G of S such that every $s \in G$ has $\text{supp}(p^s)$ consisting of $|A|$ different affinely independent beliefs. By Proposition 3, every $u \in \mathcal{U}$ and u-rationalizable $s \in R \cap G$ are such that s is uniquely u-rationalizable. Moreover, $R \cap G$ is a norm-dense subset (being the intersection of two norm-dense subsets, one of which is open) of S^κ that is also open in S^κ whenever R is (which is true whenever Ω is finite).

It remains to show the set of utilities that rationalize the SCRs in $R \cap G$ is open. But these utilities are the ones that rationalize SCRs in G (resp. $R \cap G$) if Ω is infinite (resp. finite)—a weak*-open set in either case—and so Lemma 5 applies directly. □

Proof of Proposition 4. First, observe the set S^κ_{cont} of stochastic choice rules in S^κ with a continuous version is $\|\cdot\|_\infty$-closed in S^κ. To see it is closed, note that because μ_0 has full support, the natural quotient taking a continuous function to its μ_0-a.e. equivalence class is an isometry from the space of of continuous functions $\Omega \to \mathbb{R}$ (with the supremum norm) into $L^\infty(\mu_0)$. Because the space of continuous functions is complete, its image under this isometry is necessarily closed in $L^\infty(\mu_0)$. Hence, being the intersection of S^κ with a power of this set, the set S^κ_{cont} is closed in S^κ.

Now, because S^κ_{cont} is closed, any given dense subset of S^κ is contained in S^κ_{cont} if and only if $S^\kappa_{\text{cont}} = S^\kappa$. In particular, Proposition 2 (resp. Theorem 1 under Assumption A1) implies every (uniquely) rationalizable stochastic choice rule lives in S^κ_{cont} if and only if $S^\kappa_{\text{cont}} = S^\kappa$. This observation delivers the first part of the proposition. It would also deliver the second and third parts if we could show that, assuming S^κ has a nonempty interior, every element of S^κ has a continuous version if and only if Ω is finite.

Suppose S^κ has a nonempty interior. Our goal is now to show every element of S^κ has a continuous version if and only if Ω is finite. One direction of this equivalence is trivial, because every map on a finite metric space is continuous. Toward establishing the converse, suppose Ω is infinite. Let us now observe some $s \in S^\kappa \setminus S^\kappa_{\text{cont}}$ would exist if we had some Borel $f : \Omega \to [0, 1]$ that is not μ_0-a.e. equal to a continuous function. Indeed, because S^κ has a nonempty interior in S, we know some $s^0 \in S^\kappa$ and $\epsilon > 0$ exist such that $S^\kappa \supseteq \{s \in S : |s_a - s^0_a| \leq_{\mu_0\text{-a.e.}} \epsilon, \forall a \in A\}$. Shifting s^0 and reducing ϵ if necessary, we may further assume $\epsilon \leq_{\mu_0\text{-a.e.}} s_a \leq_{\mu_0\text{-a.e.}} 1 - \epsilon$ for every $a \in A$. Then, fix two distinct $a_+, a_- \in A$, and define $s^1 \in \mathcal{S}$ by letting $s^1_{a_+} := s^0_{a_+} + \epsilon f$, $s^1_{a_-} := s^0_{a_-} - \epsilon f$, and $s^1_a := s^0_a$ for every $a \in A \setminus \{a_+, a_-\}$. By construction, $s^1 \in S^\kappa$ too, and $s^1 - s^0$ has no continuous version. Therefore, at least one of $\{s^0, s^1\}$ has no continuous version, as desired.

All that remains now is to construct (for infinite Ω) some Borel $f : \Omega \to [0, 1]$ that is not μ_0-a.e. equal to a continuous function. Because Ω is an infinite compact metrizable space, it has some sequence $\{\omega_n\}_{n=1}^\infty$ without repetition that converges to some ω_∞. In particular, ω_∞ is the sole accumulation point of $\{\omega_m\}_{m \in \mathbb{N} \cup \{\infty\}}$. Dropping to a subsequence if necessary, we may assume $\omega_n \neq \omega_\infty$ for every $n \in \mathbb{N}$. Now, take some ρ that metrizes the topology on Ω. For each $n \in \mathbb{N}$, let $r_n > 0$ be such that the $\rho(\omega_n, \omega_m) > 2r_n$ for every $m \in (\mathbb{N} \cup \{\infty\}) \setminus \{n\}$, and let N_n be the ρ-ball of radius r_n centered on ω_n. By the triangle inequality, the neighborhoods $\{N_n\}_{n=1}^\infty$ are pairwise disjoint. Moreover, that $\omega_n \to \omega_\infty$ implies $r_n \to 0$ as $n \to \infty$. Define now the bounded Borel function $f := \mathbf{1}_{\bigcup_{n=1}^\infty N_{2n}} : \Omega \to [0, 1]$, which we show has no continuous version. To do so, take an arbitrary Borel function $\tilde{f} : \Omega \to \mathbb{R}$ that is μ_0-a.e. equal to f. To conclude the proof, we show $\tilde{f}$ cannot be continuous. For any $n \in \mathbb{N}$, that μ_0 has full support implies $\mu_0(N_n) > 0$, and so some $\tilde{\omega}_n \in N_n$ has $\tilde{f}(\tilde{\omega}_n) = f(\tilde{\omega}_n)$. Then, $\rho(\tilde{\omega}_n, \omega_\infty) \leq r_n + \rho(\omega_n, \omega_\infty) \to 0$ as $n \to \infty$. However, $\tilde{f}(\tilde{\omega}_n) = f(\tilde{\omega}_n)$ is equal to 0 for odd n and 1 for even n, so that $\tilde{f}$ is discontinuous at ω_∞. $\square$

A.5. On Differentiable Costs

We begin this subsection with some convenient notations. First, given $u \in \mathcal{U}$, let

$$\begin{aligned} v_u : \Delta\Omega &\to \mathbb{R} \\ \mu &\mapsto \max_{a \in A} \int_\Omega u_a(\omega)\; \mu(\mathrm{d}\omega). \end{aligned}$$

Notation 1. *Let* $\operatorname{feas} C$ *be the (convex) set of* $p \in \mathcal{P}^F \cap \mathcal{P}^C$ *such that every* $q \in \mathcal{P}^F$ *admits some* $\epsilon \in (0,1)$ *with* $p + \epsilon(q-p) \in \mathcal{P}^C$.

Analogously, for any $c \in \mathcal{C}$, *let* $\operatorname{feas} c$ *be the (convex) set of simply drawn posteriors* μ *such that every simply drawn posterior* $\tilde{\mu}$ *admits some* $\epsilon \in (0,1)$ *with* $c\left(\mu + \epsilon(\tilde{\mu} - \mu)\right) < \infty$.

Note $\operatorname{feas} C = \mathcal{P}^F$ if C assigns a finite cost to every simple information policy; and analogously, $\operatorname{feas} c$ is the set of all simply drawn posteriors whenever c assigns a finite cost to every simply drawn posterior.

The following lemma gives an equivalent optimality condition for an SCR: the agent responds optimally to any revealed posterior, and, assuming the agent responds optimally to any hypothetical posterior, the information is optimal.

Lemma 8. *Given* $u \in \mathcal{U}$ *and* $s \in S$, *the following are equivalent:*

(i) The stochastic choice rule s *is optimal; that is,* $s \in \operatorname{argmax}_{t \in S^\kappa} \{\mathbb{E}[u \cdot t] - \kappa(t)\}$.

(ii) Every $a \in \operatorname{supp}(s)$ *has* $\int u_a(\omega)\; \mu_a^s(\mathrm{d}\omega) = v_u(\mu_a^s)$, *and*

$$p^s \in \operatorname{argmax}_{p \in \mathcal{P}^C} \left[\int v_u(\mu)\; p(\mathrm{d}\mu) - C(p)\right].$$

Proof. Let $\alpha^u : \Delta\Omega \to A$ be a measurable selection of $\mu \mapsto \operatorname{argmax}_{a \in A} \int u_a(\omega)\; \mu(\mathrm{d}\omega)$, which exists by the measurable maximum theorem (Aliprantis and Border, 2006, Theorem 18.19).

To see (i) implies (ii), we suppose (ii) does not hold. Because $v_u(\mu) \geq \int u_a(\omega)\; \mu(\mathrm{d}\omega)$ for all $a \in A$ and $\mu \in \Delta\Omega$, if $v_u(\mu_a^s) \neq \int u_a(\omega)\; \mu_a^s(\mathrm{d}\omega)$ for some $a \in \operatorname{supp}(s)$, it must be that $v_u(\mu_a^s) > \int u_a(\omega)\; \mu_a^s(\mathrm{d}\omega)$, and so the SCR t induced by (p^s, α^u) is a strict improvement over s. Similarly, if $p^s \notin \operatorname{argmax}_{p \in \mathcal{P}^C} \left[\int v_u(\mu)\; p(\mathrm{d}\mu) - C(p)\right]$, one can pick $p \in \mathcal{P}^C$ with $\int v_u(\mu)\; p(\mathrm{d}\mu) - C(p) > \int v_u(\mu)\; p^s(\mathrm{d}\mu) - C(p^s)$, in which case the SCR t induced by (p, α^u) is strictly better than s. Either way, (i) fails.

To see (ii) implies (i), suppose (ii) holds. Then, any SCR t has

$$\begin{aligned}
\mathbb{E}\left[u \cdot t\right] - \kappa(t) &= \mathbb{E}\left[u \cdot t\right] - C(p^t) \text{ (by Lemma 1)} \\
&= \left[\sum_{a \in \operatorname{supp}(p^t)} p_a^t \int u_a(\omega)\, \mu_a^t(\mathrm{d}\omega)\right] - C(p^t) \\
&\leq \left[\sum_{a \in \operatorname{supp}(p^t)} p_a^t v_u(\mu_a^t)\right] - C(p^t) \\
&= \int v_u(\mu)\, p^t(\mathrm{d}\mu) - C(p^t),
\end{aligned}$$

where the inequality holds with equality for $t = s$ by hypothesis. Then, s maximizes $t \mapsto \mathbb{E}\left[u \cdot t\right] - \kappa(t)$ because p^s maximizes $p \mapsto \int v_u(\mu)\, p(\mathrm{d}\mu) - C(p)$. □

Our definition of a derivative of C assumed the derivative c was convex, which made posterior-separable approximation C_c a valid cost function by Jensen's inequality. Here, we note that under sufficient regularity, this convexity property is redundant, being implied by monotonicity of C. See Appendix C for the proof.

Fact 1. *Let $p \in \operatorname{feas} C$, and let $c : \Delta\Omega \to \mathbb{R}$ be a continuous function. If every $p' \in \mathcal{P}^C$ has*

$$d_p^+ C(p') = \int c(\mu)\, (p' - p)(\mathrm{d}\mu),$$

then c is convex. In particular, in this case, $c \in \mathcal{C}$, so that c is a derivative of C at p.

The following lemma establishes an equivalence between optimal information choice for the agent's cost function and optimal information choice for a posterior-separable approximation of her cost function.

Lemma 9. *If c is a derivative of C at $p \in \operatorname{feas} C$, then any $u \in \mathcal{U}$ has*

$$p \in \operatorname{argmax}_{q \in \mathcal{P}} \left[\int v_u(\mu)\, q(\mathrm{d}\mu) - C(q)\right] \iff p \in \operatorname{argmax}_{q \in \mathcal{P}} \int \left[v_u(\mu) - c(\mu)\right]\, q(\mathrm{d}\mu).$$

Proof. Suppose first $p \in \operatorname{argmax}_{q \in \mathcal{P}} \int \left[v_u(\mu) - c(\mu)\right]\, q(\mathrm{d}\mu)$. Then, for every $q \in \mathcal{P}^C$,

$$\int v_u(\mu)\, (q - p)(\mathrm{d}\mu) \leq \int c(\mu)\, (q - p)(\mathrm{d}\mu) \leq C(q) - C(p),$$

where the last inequality follows from $C\left(p + \epsilon(q - p)\right)$ being convex in $\epsilon \in [0, 1]$. Because $C(q) = \infty$ for all $q \in \mathcal{P} \setminus \mathcal{P}^C$, the left-hand-side condition follows.

Conversely, suppose some $\tilde{p} \in \mathcal{P}$ has $\int [v_u(\mu) - c(\mu)]\ \tilde{p}(\mathrm{d}\mu) > \int [v_u(\mu) - c(\mu)]\ p(\mathrm{d}\mu)$. Because c is convex and lower semicontinuous, and v_u is the maximum of finitely many affine functions, we may assume without loss that $\tilde{p} \in \mathcal{P}^F$. Observe every $\epsilon \in (0,1)$ satisfies

$$\int [v_u(\mu) - c(\mu)]\ [p + \epsilon(\tilde{p} - p)]\,(\mathrm{d}\mu) > \int [v_u(\mu) - c(\mu)]\ p(\mathrm{d}\mu).$$

Thus, because $p \in \operatorname{feas} C$, we may assume without loss some convex combination p' of $\tilde{p}$ and p has $p^{'} \in \mathcal{P}^C$. But then

$$\begin{aligned}
0 &< \int v_u(\mu)\ (p' - p)(\mathrm{d}\mu) - \int c(\mu)\ (p' - p)(\mathrm{d}\mu) \\
&= \lim_{\epsilon \searrow 0} \frac{1}{\epsilon} \left\{ \int v_u(\mu)\ [p + \epsilon(p' - p)]\,(\mathrm{d}\mu) - C\left(p + \epsilon(p' - p)\right) - \int v_u(\mu)\, p(\mathrm{d}\mu) + C(p) \right\},
\end{aligned}$$

so that small enough $\epsilon \in (0,1)$ will satisfy

$$\int v_u(\mu)\ [p + \epsilon(p' - p)]\,(\mathrm{d}\mu) - C\left(p + \epsilon(p' - p)\right) > \int v_u(\mu)\, p(\mathrm{d}\mu) - C(p).$$

The lemma follows. □

The next lemma establishes an equivalence between optimality of an SCR for our agent's cost function and its optimality for a posterior-separable approximation of the same.

Lemma 10. *Suppose $s \in S$ has $p^s \in \operatorname{feas} C$ and c is a derivative of C at p^s. Given $u \in \mathcal{U}$,*

$$s \in \operatorname{argmax}_{s' \in S} \left[\mathbb{E}\left[u \cdot s'\right] - \kappa(s')\right] \iff s \in \operatorname{argmax}_{s' \in S} \left[\mathbb{E}\left[u \cdot s'\right] - \kappa_c(s')\right].$$

Proof. Lemma 8 says s is optimal with cost κ if and only if $\int u_a(\omega)\ \mu^s_a(\mathrm{d}\omega) = v_u(\mu^s_a)$ for all $a \in \operatorname{supp}(s)$ and $p^s \in \operatorname{argmax}_{p \in \mathcal{P}^C} \left[\int v_u(\omega)\, p(\mathrm{d}\omega) - C(p)\right]$. Further, the information cost C_c satisfies our standing hypotheses on C. Lemma 8 therefore tells us s is optimal with cost κ_c if and only if $\int u_a(\omega)\ \mu^s_a(\mathrm{d}\omega) = v_u(\mu^s_a)$ for all $a \in \operatorname{supp}(s)$ and $p^s \in \operatorname{argmax}_{q \in \mathcal{P}^C} \left[\int v_u(\mu)\, q(\mathrm{d}\mu) - C_c(q)\right]$. The equivalence would therefore follow if we knew $p^s \in \operatorname{argmax}_{p \in \mathcal{P}^C} \left[\int v_u(\mu)\, p(\mathrm{d}\mu) - C(p)\right]$ if and only if

$$p^s \in \operatorname{argmax}_{p \in \mathcal{P}^C} \left[\int v_u(\mu)\, p(\mathrm{d}\mu) - C_c(p)\right],$$

which Lemma 9 guarantees. □

The next lemma gives an explicit formula for the directional derivatives of the indirect cost function when information costs are posterior separable.

Lemma 11. *If* $c \in \mathcal{C}$ *and* $s, s' \in S$ *are such that every* $\mathrm{supp}(p^s) \subseteq \mathrm{feas}\ c$*, then*[33]

$$\begin{aligned} d_s^+ \kappa_c(s') &= \sum_{a \in \mathrm{supp}(s)} \left[(p_a^{s'} - p_a^s)\, c(\mu_a^s) + p_a^{s'}\, d^+_{\mu_a^s} c(\mu_a^{s'}) \right] \\ &\quad + \sum_{a \in \mathrm{supp}(s') \setminus \mathrm{supp}(s)} p_a^{s'}\, c(\mu_a^{s'}). \end{aligned}$$

Proof. Let $\hat{A} := \mathrm{supp}(s) \cup \mathrm{supp}(s')$, and define $s^\epsilon := s + \epsilon(s' - s)$ for each $\epsilon \in (0,1)$. Any $\epsilon \in (0,1)$ has $\mathrm{supp}(s^\epsilon) = \hat{A}$ and, for each $a \in \hat{A}$,

$$\begin{aligned} p_a^{s^\epsilon} &= p_a^s + \epsilon(p_a^{s'} - p_a^s) \\ \mu_a^{s^\epsilon} &= \mu_a^s + \epsilon \tfrac{p_a^{s'}}{p_a^{s^\epsilon}} (\mu_a^{s'} - \mu_a^s). \end{aligned}$$

Therefore, every $\epsilon \in (0,1)$ and $a \in \hat{A}$ have

$$\eta_a(\epsilon) := \tfrac{1}{\epsilon} \left[p_a^{s^\epsilon}\, c(\mu_a^{s^\epsilon}) - p_a^s\, c(\mu_a^s) \right] = (p_a^{s'} - p_a^s)\, c(\mu_a^{s^\epsilon}) + p_a^s\, \tfrac{1}{\epsilon} \left[c(\mu_a^{s^\epsilon}) - c(\mu_a^s) \right].$$

Hence, $\eta_a(\epsilon)$ is equal to $p_a^{s'} c(\mu_a^{s'})$ if $p_a^s = 0$, is equal to $-p_a^s c(\mu_a^s)$ if $p_a^{s'} = 0$, and is otherwise equal to

$$(p_a^{s'} - p_a^s)\, c(\mu_a^{s^\epsilon}) + p_a^{s'}\, \frac{p_a^s}{p_a^{s^\epsilon}}\, \frac{1}{\epsilon \frac{p_a^{s'}}{p_a^{s^\epsilon}}} \left[c(\mu_a^{s^\epsilon}) - c(\mu_a^s) \right],$$

which converges as $\epsilon \to 0$ (because c is convex, and hence continuous on any open line segment on its domain, and $\mathrm{supp}(p^s) \subseteq \mathrm{feas}\ c$) to

$$\begin{aligned} & (p_a^{s'} - p_a^s)\, c\left(\lim_{\epsilon \searrow 0} \mu_a^{s^\epsilon} \right) + p_a^{s'}\, \frac{p_a^s}{\lim_{\epsilon \searrow 0} p_a^{s^\epsilon}}\ \lim_{\tilde{\epsilon} \searrow 0} \frac{1}{\tilde{\epsilon}} \left[c(\mu^{s + \tilde{\epsilon}(s' - s)}) - c(\mu_a^s) \right] \\ = \ & (p_a^{s'} - p_a^s)\, c(\mu_a^s) + p_a^{s'}\ 1\ d^+_{\mu_a^s} c(\mu_a^{s'}). \end{aligned}$$

Hence,

$$\begin{aligned} \lim_{\epsilon \searrow 0} \tfrac{1}{\epsilon} \left[\kappa_c(s + \epsilon(s' - s)) - \kappa_c(s) \right] &= \lim_{\epsilon \searrow 0} \tfrac{1}{\epsilon} \int c \mathrm{d}(p^{s^\epsilon} - p^s) = \lim_{\epsilon \searrow 0} \sum_{a \in \hat{A}} \eta_a(\epsilon) \\ &= \sum_{a \in \mathrm{supp}(s)} \left[(p_a^{s'} - p_a^s)\, c(\mu_a^s) + p_a^{s'}\, d^+_{\mu_a^s} c(\mu_a^{s'}) \right] \\ &\quad + \sum_{a \in \mathrm{supp}(s') \setminus \mathrm{supp}(s)} p_a^{s'}\, c(\mu_a^{s'}), \end{aligned}$$

[33] When $p_a^{s'} = 0$, we adopt the convention that $p_a^{s'}\ d^+_{\mu_a^s} c(\mu_a^{s'})$ is zero too.

as desired. $\square$

The following lemma shows derivatives preserve the finite-cost property of any information policy.

Lemma 12. *If c is a derivative of C at $p \in \mathcal{P}^C$, then $\int c(\mu)\, p'(\mathrm{d}\mu) < \infty$ for every $p' \in \mathcal{P}^C$.*

Proof. Consider any $p' \in \mathcal{P}^C$. Convexity of C implies $d_p^+ C(p') \in \mathbb{R} \cup \{-\infty\}$. Hence, applying the definition of a derivative (including that $\int c(\mu)\, p(\mathrm{d}\mu) < \infty$),

$$\int c(\mu)\, p'(\mathrm{d}\mu) = \int c(\mu)\, p(\mathrm{d}\mu) + d_p^+ C(p') < \infty,$$

as required. $\square$

A.6. On Iteratively Differentiable Costs

Although not relevant to our subsequent results, let us briefly note a uniqueness property (proven in Appendix C) for iterated differentiability that justifies the notation ∇c_μ.

Fact 2. *Any function in $\mathcal{C}$ admits at most one derivative at a simply drawn posterior.*

The following lemma yields a more explicit form (relative to Lemma 11) for the directional derivative of a posterior-separable approximation of the cost function, in the iteratively differentiable case.

To state a slightly more general version of the result (which will be helpful for one result described in section 7), we invest in another definition. For any $c \in \mathcal{C}$ and any simply drawn posterior μ, define the **subdifferential**

$$\partial c(\mu) := \left\{ f \in L^1(\mu_0) : \ c(\mu') \geq c(\mu) + \int f(\omega)\, (\mu' - \mu)(\mathrm{d}\omega)\ \forall \text{ simply drawn } \mu' \in \Delta\Omega \right\}$$

of c at μ. Clearly, if c is differentiable at μ, then $\nabla c_\mu \in \partial c(\mu)$ because c is convex.

Remark 1. *Although we adopt the notation ∂c for parsimony, the above definition is best understood as the subdifferential of a function $\tilde{c} : L^\infty(\mu_0) \to \overline{\mathbb{R}}$, where we identify each element of $L^\infty(\mu_0)$ with the measure on Ω whose Radon-Nikodym derivative with respect to μ_0 is given by that element.*

Lemma 13. *Let $c \in \mathcal{C}$ and $s, s' \in S$ have $\mathrm{supp}(p^s) \subseteq \mathrm{feas}\ c$, and let $f_a \in \partial c(\mu_a^s)$ have $\int f_a(\omega)\ \mu_a^s(\mathrm{d}\omega)$ for each $a \in \mathrm{supp}(s)$. Let $u^{s,s'} \in \mathcal{U}$ have $u_a^{s,s'} = f_a$ for $a \in \mathrm{supp}(s)$ and*

$u_a^{s,s'} = c(\mu_a^{s'})\mathbf{1}$ *for* $a \in \operatorname{supp}(s') \setminus \operatorname{supp}(s)$. *Then,*

$$d_s^+\kappa_c(s') \geq \mathbb{E}\left[(s'-s)\cdot u\right],$$

with equality holding if $f_a = \nabla c_{\mu_a^s}$ *for every* $a \in \operatorname{supp}(s)$.

Proof. Let $u := u^{s,s'}$. Then, every $a \in \operatorname{supp}(s)$ has

$$\begin{aligned}
p_a^{s'}\, d_{\mu_a^s}^+ c(\mu_a^{s'}) &\geq p_a^{s'} \int f_a(\omega)\,(\mu_a^{s'} - \mu_a^s)(\mathrm{d}\omega) \\
&= p_a^{s'} \int \left[\frac{s_a'(\omega)}{p_a^{s'}} - \frac{s_a(\omega)}{p_a^s}\right] f_a(\omega)\,\mu_0(\mathrm{d}\omega) \\
&= \int \left[s_a'(\omega) - s_a(\omega)\right] f_a(\omega)\,\mu_0(\mathrm{d}\omega) + \int \left(p_a^s - p_a^{s'}\right) \frac{s_a(\omega)}{p_a^s} f_a(\omega)\,\mu_0(\mathrm{d}\omega) \\
&= \int \left[s_a'(\omega) - s_a(\omega)\right] u(\omega)\,\mu_0(\mathrm{d}\omega) - \left(p_a^{s'} - p_a^s\right) \int f_a(\omega)\,\mu_a^s(\mathrm{d}\omega) \\
&= \mathbb{E}\left[(s_a' - s_a)u_a\right] - (p_a^{s'} - p_a^s)\, c(\mu_a^s).
\end{aligned}$$

Moreover, the above inequality holds with equality if $f_a = \nabla c_{\mu_a^s}$ for every $a \in \operatorname{supp}(s)$. Therefore, Lemma 11 implies

$$\begin{aligned}
d_s^+\kappa_c(s') &\geq \sum_{a\in\operatorname{supp}(s)} \mathbb{E}\left[(s_a' - s_a)u_a\right] + \sum_{a\in\operatorname{supp}(s')\setminus\operatorname{supp}(s)} p_a^{s'}\, c(\mu_a^{s'}) \\
&= \sum_{a\in\operatorname{supp}(s)} \mathbb{E}\left[(s_a' - s_a)u_a\right] + \sum_{a\in\operatorname{supp}(s')\setminus\operatorname{supp}(s)} \mathbb{E}\left[(s_a' - s_a)\, c(\mu_a^{s'})\right] \\
&= \mathbb{E}\left[(s'-s)\cdot u\right],
\end{aligned}$$

again with equality if $f_a = \nabla c_{\mu_a^s}$ for every $a \in \operatorname{supp}(s)$. The result follows. □

The next lemma relates the set $\operatorname{feas} C$ to the corresponding set of beliefs for its derivatives.

Lemma 14. *If* c *is a derivative of* C *at* $p \in \operatorname{feas} C$*, then* $\operatorname{supp}(p) \subseteq \operatorname{feas} c$.

Proof. Consider $\mu \in \operatorname{supp}(p)$, and any simply drawn posterior μ'; we must show some proper convex combination of μ and μ' belongs to $c^{-1}(\mathbb{R})$.

By hypothesis, p is simple, $\mu \in \operatorname{supp}(p)$, and μ' is simply drawn. Hence, some finite-support $q, q' \in \Delta\Delta\Omega$ and $\beta, \beta' \in (0,1]$ exist with $p = (1-\beta)q + \beta\delta_\mu$ and $p' := (1-\beta')q' + \beta'\delta_{\mu'} \in \mathcal{P}^C$. Now, that $p \in \operatorname{feas} C$ implies some $\epsilon \in (0,1)$ has $(1-\epsilon)p + \epsilon p' \in \mathcal{P}^C$. Define

$$\mu^\epsilon := \tfrac{(1-\epsilon)\beta}{(1-\epsilon)\beta+\epsilon\beta'}\mu + \tfrac{\epsilon\beta'}{(1-\epsilon)\beta+\epsilon\beta'}\mu',$$

and

$$p^\epsilon := (1-\epsilon)(1-\beta)q + \epsilon(1-\beta')q' + [(1-\epsilon)\beta + \epsilon\beta']\,\delta_{\mu^\epsilon}.$$

Because C is monotone and $p^\epsilon \preceq (1-\epsilon)p + \epsilon p'$ by construction, $p^\epsilon \in \mathcal{P}^C$. Hence, Lemma 12 tells us $\int c(\mu)\, p^\epsilon(\mathrm{d}\mu) < \infty$, and so too $c(\mu^\epsilon) < \infty$. The lemma follows. ◻

The following lemma gives an exact optimality condition for a given SCR when information costs are iteratively differentiable.

Lemma 15. *Let $s \in S$ have $p^s \in \operatorname{feas} C$, suppose C is iteratively differentiable at p^s with derivative c. For $u \in \mathcal{U}$, the following are equivalent:*

(i) SCR s is u-rationalizable.

(ii) Every $s' \in S$ has $\mathbb{E}\left[(u^{s,s'} - u)\cdot(s'-s)\right] \geq 0$, where $u^{s,s'} \in \mathcal{U}$ has $u_a^{s,s'} = \nabla c_{\mu_a^s}$ for $a \in \operatorname{supp}(s)$ and $u_a^{s,s'} = c(\mu_a^{s'})\mathbf{1}$ for $a \in \operatorname{supp}(s') \setminus \operatorname{supp}(s)$

Proof. By Lemma 10, u rationalizes s if and only if u would rationalize s given alternative information cost C_c. Hence, Lemma 4 (applied to the model with cost C_c) tells us u rationalizes s if and only if every $s' \in S^\kappa$ has $d_s^+\kappa_c(s') \geq \mathbb{E}\left[u\cdot(s'-s)\right]$. Because Lemma 14 tells us $\operatorname{supp}(p^s) \subseteq \operatorname{feas} c$, Lemma 13 shows $d_s^+\kappa_c(s') = \mathbb{E}\left[u^{s,s'}\cdot(s'-s)\right]$ for every $s' \in S$. The equivalence follows. ◻

The following lemma characterizes the utilities that can rationalize a given full-support SCR when information costs are iteratively differentiable. Letting c denote a derivative of C at the SCR, all such utilities are given by the derivative of c at the corresponding revealed posterior, augmented by a nuisance term, potentially with an additional payoff penalty for actions not chosen in a given state.

Lemma 16. *Let $s \in S$ have $p^s \in \operatorname{feas} C$ and $\operatorname{supp}(s) = A$. Suppose C is iteratively differentiable at p^s with derivative c. The following are equivalent for $u \in \mathcal{U}$:*

(i) SCR s is u-rationalizable.

(ii) Some $\lambda \in L^1(\mu_0)$ and $\gamma \in L^1(\mu_0)_+^A$ exist such that every $a \in A$ has

$$\begin{aligned} u_a &= \lambda - \gamma_a + \nabla c_{\mu_a^s}, \\ s_a\gamma_a &= 0. \end{aligned}$$

Proof. Let $u^s := u - (\nabla c_{\mu_a^s})_{a\in A} \in \mathcal{U}$. By Lemma 15, u rationalizes s if and only if every $s' \in S$ has $\mathbb{E}\left[-u^s \cdot (s' - s)\right] \geq 0$, or equivalently, $s \in \text{argmax}_{s'\in S}\mathbb{E}\left[u^s \cdot s'\right]$. Hence, u rationalizes s if and only if μ_0-almost every ω has

$$\{a \in A :\ s_a(\omega) > 0\} \subseteq \text{argmax}_{a\in A} u_a^s(\omega).$$

But the latter condition holds if and only if some $\lambda \in L^1(\mu_0)$ and $\gamma \in L^1(\mu_0)_+^A$ have $u^s = (\lambda - \gamma_a)_{a\in A}$ and $(\gamma_a s_a)_{a\in A} = 0$: The "if" direction is immediate, and the "only if" direction comes from considering $\lambda := \max_{a\in A} u_a^s$. The lemma follows. □

A.7. Section 6 Proofs

Proof of Theorem 2. Because C is finite on $\mathcal{P}^F$, every SCR s has $p^s \in \text{feas}\, C$. The theorem therefore follows directly from Lemma 10. □

Proof of Proposition 5. Because C is finite on $\mathcal{P}^F$, every SCR s has $p^s \in \text{feas}\, C$. The proposition therefore follows directly from Lemma 16. □

Let us briefly note the infinite-slope condition, Assumption A2(iii)—which says costs decrease infinitely steeply as one moves from a not-fully-mixed information policy toward providing no information—could be equivalently replaced with a more permissive unbounded steepness condition. See Appendix C for the proof.

Fact 3. *For any $p \in \mathcal{P}^C$, the following are equivalent:*

(i) $d_p^+ C(\delta_{\mu_0}) = -\infty$.

(ii) $\inf_{q\in\mathcal{P}^C,\ \epsilon\in(0,1)} \frac{1}{\epsilon}\left[C(p + \epsilon(q - p)) - C(p)\right] = -\infty$.

The following corollary follows immediately from Lemma 16 (and the weaker version requiring C to be finite on $\mathcal{P}^F$ follows immediately from Proposition 5).

Corollary 3. *Let $s \in S$ have $p^s \in \text{feas}\, C$ and s have conditionally full support. If C is iteratively differentiable at p^s, and u and u' both rationalize s, some $\lambda \in L^1(\mu_0)$ exists such that*

$$u_a = u'_a + \lambda\ \forall a \in A.$$

Next, we show Assumption A2(iii) means an optimal action recommendation from s never rules out any state.

Lemma 17. *Suppose C satisfies Assumption A2(iii). If s is rationalizable, then any $a \in \operatorname{supp} s$ has s_a strictly positive μ_0-almost surely.*

Proof. Take $u \in \mathcal{U}$, and suppose $s \in S$ admits some $a \in \operatorname{supp} s$ such that s_a is not μ_0-almost surely strictly positive. We wish to show s is not u-rationalizable.

We have nothing to show if $s \notin S^\kappa$, so we focus on the case in which $s \in S^\kappa$. Pick some $a_0 \in A$ and let t be the unique SCR with $t_{a_0} = \mathbf{1}$. Clearly, $p^t = \delta_{\mu_0}$, and so $t \in S^\kappa$ and $p^t \in \mathcal{P}^C$. For every $\epsilon \in (0,1)$, let $t^\epsilon = s + \epsilon(t - s)$. Then, because p^s is not fully mixed,

$$\begin{aligned}
\frac{1}{\epsilon}[\kappa(t^\epsilon) - \kappa(s)] &= \frac{1}{\epsilon}[C(p^{t^\epsilon}) - C(p^s)] \\
&\leq \frac{1}{\epsilon}[C((1-\epsilon)p^s + \epsilon p^t) - C(p^s)] \\
&= \frac{1}{\epsilon}[C((1-\epsilon)p^s + \epsilon\delta_{\mu_0}) - C(p^s)] \xrightarrow{\epsilon \searrow 0} -\infty,
\end{aligned}$$

where the inequality comes from monotonicity of C and Lemma 2(ii), and the limit calculation comes from Assumption A2(iii). Hence, every $u \in \mathcal{U}$ has

$$\frac{1}{\epsilon}\left\{\mathbb{E}\left[u \cdot t^\epsilon\right] - \kappa(t^\epsilon) - \left[\mathbb{E}\left[u \cdot s\right] - \kappa(s)\right]\right\} = \mathbb{E}\left[u \cdot (t - s)\right] - \frac{1}{\epsilon}\left[\kappa(t^\epsilon) - \kappa(s)\right] \xrightarrow{\epsilon \searrow 0} -\infty.$$

Thus, for all sufficiently small $\epsilon \in (0,1)$, the agent's objective must be strictly higher under t^ϵ than under s. That is, s is not rationalizable. ◻

The following lemma shows a norm-dense set of uniquely rationalizable SCRs can be leveraged to find a weak*-dense set of SCRs generating unique subset predictions.

Lemma 18. *Suppose Assumption A2 holds and some norm-dense subset S_1 of S^κ exists that comprises only uniquely rationalizable SCRs. Then, the set of SCRs yielding unique subset predictions is weak* dense in S.*

Proof. By Proposition 1, every $B \in \mathcal{A}$ admits a co-meager set of utilities $\mathcal{U}_B \subseteq \mathcal{U}$ that generate a unique prediction over B; that is, each $u \in \mathcal{U}_B$ admits a unique s that is u-rationalizable over B. It follows the set $\mathcal{U}_\mathcal{A} = \cap_{B\in\mathcal{A}}\mathcal{U}_B$ is a co-meager (and therefore dense) subset of $\mathcal{U}$ such that, for every B, every $u \in \mathcal{U}_\mathcal{A}$ uniquely rationalizes some $s \in S^B$ over B.

Observe now, the set S_1 is norm-dense in S because Assumption A2(i) implies $S^\kappa = S$. Let S_0 denote the set of full-support SCRs, which is weak* open (hence norm open) and norm dense in S. Therefore, $\tilde{S} := S_0 \cap S_1$ is norm dense in S too. Because every norm-dense set is weak* dense, it suffices to show $\tilde{S}$ is contained in the weak* closure of the SCRs generating unique subset predictions. To that end, take any $s \in \tilde{S}$, and let $\mathcal{N} \subseteq \mathcal{S}$ be a weak* neighborhood of s; we want to show some SCR in $\mathcal{N}$ generates unique subset predictions.

Let u be a utility that uniquely rationalizes s (over A). Some sequence $(u^n)_{n\in\mathbb{N}}$ from $\mathcal{U}_{\mathcal{A}}$ converges to u because $\mathcal{U}_{\mathcal{A}}$ is dense in $\mathcal{U}$. Let s^n denote the SCR rationalized by u^n for each $n \in \mathbb{N}$. Because $s^n \in \partial V(u^n)$ by Lemma 4, and ∂V norm-to-weak* upper hemicontinuous (Lemma 3) and single valued at u, it follows that $s^n \xrightarrow{w^*} s$ as $n \to \infty$. In particular, $s^n \in \mathcal{N}$ for all n sufficiently large.

Next, observe s^n has full support (i.e., belongs to S_0) for large enough n, because S_0 is weak* open in S, and the weak*-limit s has full support. But Lemma 17 tells us any rationalizable full-support SCR has conditionally full support in light of Assumption A2(iii). Therefore, s^n has conditionally full support for sufficiently large n.

Now, fix $n \in \mathbb{N}$ large enough that $s^n \in \mathcal{N}$ and s^n has conditionally full support; the theorem will follow if we can show s^n generates unique subset predictions. To that end, consider any $\hat{u}$ that rationalizes s^n. Because s^n has conditionally full support, p^{s^n} is fully mixed, and so C is iteratively differentiable at p^{s^n} by Assumption A2(ii). Moreover, Assumption A2(i) implies $\operatorname{feas} C \supseteq \mathcal{P}^F \ni p^{s^n}$, and so Corollary 3 delivers some $\lambda \in L^1(\mu_0)$ for which every $a \in A$ has $\hat{u}_a = u^n_a + \lambda$. It follows that every $t \in S^\kappa$ has

$$\mathbb{E}\left[\hat{u} \cdot t\right] - \kappa(t) = \mathbb{E}\left[u^n \cdot t\right] - \kappa(t) + \mathbb{E}\left[\lambda\right],$$

and so every $B \in \mathcal{A}$ has

$$\operatorname{argmax}_{t \in S_B}\left[\mathbb{E}\left[\hat{u} \cdot t\right] - \kappa(t)\right] = \operatorname{argmax}_{t \in S_B}\left[\mathbb{E}\left[u^n \cdot t\right] - \kappa(t)\right].$$

Hence, over any B, the utility $\hat{u}$ rationalizes the same set of SCRs as u^n does. But $u^n \in \mathcal{U}_{\mathcal{A}}$, meaning it rationalizes a unique SCR over B. It follows s^n yields unique subset predictions, as required. □

Proof of Theorem 3. Theorem 1 delivers a norm-dense subset of S^κ comprising only uniquely rationalizable SCRs. The result then follows directly from Lemma 18. □

B. Online Appendix: Supplement to Section 7

This appendix provides formal support for any nontrivial claims made in section 7.

B.1. Unique Rationalizability and Strict Convexity

The following lemma shows that, when C is strictly convex, the agent's optimal information choice is unique, and the only scope for multiple best responses is the willingness to mix over her action conditional on a realized signal.

Lemma 19. *Suppose C is strictly convex, and let $u \in \mathcal{U}$. If $s, t \in S$ are u-rationalizable, then $p^s = p^t$ and, for every $a \in A$, the functions s_a and t_a are proportional.*

Proof. First, we show no two u-rationalizable SCRs can generate different information policies. To that end, suppose $s^1, s^2 \in S^\kappa$ have $p^{s^1} \neq p^{s^2}$; we show they cannot both be u-rationalizable. Letting $s := \frac{1}{2}\sum_{i=1}^2 s^i$, Lemma 2(ii) tells us $\frac{1}{2}\sum_{i=1}^2 p^{s^i} \succeq p^s$, so that

$$\kappa\left(s\right) = C\left(p^s\right) \leq C\left(\frac{1}{2}\sum_{i=1}^2 p^{s^i}\right) < \frac{1}{2}\sum_{i=1}^2 C\left(p^{s^i}\right) = \frac{1}{2}\sum_{i=1}^2 \kappa\left(s^i\right).$$

Hence, $\mathbb{E}\left[u \cdot s\right] - \kappa(s) > \frac{1}{2}\sum_{i=1}^2 \left[\mathbb{E}\left[u \cdot s^i\right] - \kappa(s^i)\right] \geq \min_{i=1,2}\left[\mathbb{E}\left[u \cdot s^i\right] - \kappa(s^i)\right]$. Therefore, s witnesses that at least one of $\{s^1, s^2\}$ is not u-rationalizable.

Now, let $s, t \in S^\kappa$ be u-rationalizable. Because $r \mapsto \mathbb{E}\left[u \cdot r\right] - \kappa(r)$ is concave (by Corollary 1), it follows that $r := \frac{1}{2}(s+t)$ is u-rationalizable as well. By the above analysis, $p^s = p^t = p^r$; in particular, $p^r \not\succ \frac{1}{2}p^s + \frac{1}{2}p^t$. Lemma 2(iii) then implies no $a \in \operatorname{supp}(s) \cap \operatorname{supp}(t)$ exists such that $\mu_a^s \neq \mu_a^t$. Said differently, s_a and t_a are proportional for every $a \in A$. $\square$

Proof of Proposition 6. Let S^1 denote the set of SCRs $s \in S^\kappa$ with $\operatorname{supp}(s) = A$ and the $|A|$ beliefs $\{\mu_a^s\}_{a \in A}$ all distinct. Below, we show S^1 is weak* open and norm dense in S^κ. Before doing so, let us see this result would deliver the proposition. To that end, let R denote the set of rationalizable SCRs if Ω is infinite, and let it denote the relative interior of S^κ if Ω is finite. Given Proposition 2, every SCR in R is rationalizable, and R is norm dense in S^κ. Moreover, by construction, R is open in S^κ if Ω is finite. Hence, the intersection $R \cap S^1$ is norm dense in S^κ and open in it if Ω is finite. Moreover, if we establish that any $u \in \mathcal{U}$ and u-rationalizable $s \in R \cap S^1$ are such that s is uniquely u-rationalizable, Lemma 5 would tell us the SCRs in $R \cap S^1$ are rationalized by an open set of utilities.[34] Finally, let us see that any $u \in \mathcal{U}$ and u-rationalizable $s \in R \cap S^1$ are such that s is uniquely u-rationalizable. To do so, take any u-rationalizable SCR t. That $s \in S^1$ implies p^s has support size $|A|$. By Lemma 19, we know $p^t = p^s$, and so p^t has support size $|A|$ too. Hence, $\text{supp}(s) = \text{supp}(t) = A$. Lemma 19 then also tells us $\mu_a^s = \mu_a^t$ for every $a \in A$. But then, because $\{\mu_a^s\}_{a \in A}$ are $|A|$ distinct beliefs, it follows from $p^s = p^t$ that $p_a^s = p_a^t$ for every $a \in A$. Hence, $t = s$, as desired.

So all that remains is to show S^1 is weak* open and norm dense in S^κ. To that end, define the set

$$S^+ := \bigcap_{a \in A} \{s \in S^\kappa : \mathbb{E}[s_a] > 0\}$$

and, for each distinct $a, b \in A$, the set

$$S^{a,b} := \bigcup_{\hat{\Omega} \subseteq \Omega \text{ Borel}} \{s \in S^\kappa : \mathbb{E}[s_a]\, \mathbb{E}[\mathbf{1}_{\hat{\Omega}} s_b] \neq \mathbb{E}[s_b]\, \mathbb{E}[\mathbf{1}_{\hat{\Omega}} s_a]\},$$

By construction, these sets are all weak* open (hence, norm open) in S^κ, and so too is the finite intersection $S^1 = S^+ \cap \bigcap_{a,b \in A \text{ distinct}} S^{a,b}$. Moreover, because a finite intersection of open and dense sets is dense, norm denseness of S^1 will follow if we can show S^+ is norm dense in S^κ and each pair of distinct $a, b \in A$ has $S^+ \cap S^{a,b}$ norm dense in S^+. To see S^+ is norm dense, note $s^+ := (\frac{1}{|A|}\mathbf{1})_{a \in A} \in S^\kappa$ because $p^{s^+} = \delta_{\mu_0}$ and $C(\delta_{\mu_0}) < \infty$. Because S^κ is convex by Corollary 1, it follows that every $s \in S^\kappa$ and $\epsilon \in (0,1)$ have $(1-\epsilon)s + \epsilon s^+ \in S^\kappa$, which witnesses (taking $\epsilon \to 0$) the SCR s as a norm limit from S^+.

Now, fix any pair of distinct $a, b \in A$. It remains to show $S^+ \cap S^{a,b}$ is norm dense in S^+. To that end, let $s \in S^+$ be arbitrary; we want to show s is in the norm closure of $S^+ \cap S^{a,b}$. If $s \in S^{a,b}$, we have nothing to show, so we focus on the complementary case in which $\mu_a^s = \mu_b^s =: \mu$. Below, we locate an SCR $s^* \in S^\kappa$ such that any proper convex combination of s and s^* lies in $S^{a,b}$. Observe such proper convex combinations necessarily live in S^κ (because κ is convex) and so in S^+ (because); and they can approximate s arbitrarily well by choosing sufficiently skewed weights. Hence, finding such an s^* will yield the required denseness property. To locate such an s^*, we separately address the case in which $\mu \neq \mu_0$ and the case in which $\mu = \mu_0$.

First, suppose $\mu \neq \mu_0$, and let s^* denote the unique SCR with $s_a^* = \mathbf{1}$. That $C(\delta_{\mu_0}) < \infty$ implies $s^* \in S^\kappa$. Moreover, any proper convex combination t of s and s^* is in $S^{a,b}$ because it

[34] Just as in the proof of Theorem 1, we can apply Lemma 5 to the weak*-open set G (resp. $R \cap G$) if Ω is infinite (resp. finite).

has $t_b = \mu$, whereas t_a is a proper convex combination of μ and μ_0. Thus, s^* is as required.

Finally, suppose $\mu = \mu_0$. By hypothesis, some $p \in \mathcal{P}^C$ exists such that $p \neq \delta_{\mu_0}$. Because p has barycenter μ_0, the distribution p must be nondegenerate. Pooling different posteriors generated by p if necessary (which will weakly reduce costs and so remain in $\mathcal{P}^C$ because C is monotone), we may assume without loss p has binary support $\{\mu_a^*, \mu_b^*\}$. Note μ_0 lies strictly between μ_a^* and μ_b^*. Then, define s^* to be the unique SCR with $\mu_a^{s^*} = \mu_a^*$ and $\mu_b^{s^*} = \mu_b^*$. By construction, $\kappa(s^*) = C(p^{s^*}) = C(p) < \infty$; that is, $s^* \in S^\kappa$. Finally, any proper convex combination t of s and s^* is in $S^{a,b}$ because it has t_a and t_b being proper convex combinations of μ_0 with μ_a^* and μ_b^*, respectively. Thus, s^* is as required. □

Proof of Proposition 7. Assumption A2(i) implies, given $|\Omega| > 1$, that $\mathcal{P}^C \neq \{\delta_{\mu_0}\}$. Hence, Proposition 6 delivers a norm-dense subset of S^κ comprising only uniquely rationalizable SCRs. The result then follows directly from Lemma 18. □

B.2. Subdifferentiability, Rationalizability, and Posterior Separable Costs

Proposition 10. *Fix some $s \in S$ such that $p^s \in \operatorname{feas} C$, and suppose c is a derivative of C at p^s. If $\partial c(\mu_a^s)$ is nonempty for all $a \in \operatorname{supp} s$, then $\partial \kappa(s)$ is nonempty, and so s is rationalizable.*

Proof. By Lemma 10, it suffices to show s is rationalizable according to κ_c, which is (by Lemma 4) equivalent to some $u \in \mathcal{U}$ satisfying, for every $s' \in S^\kappa$, the inequality $d_s^+ \kappa_c(s') \geq \mathbb{E}\left[u \cdot (s' - s)\right]$.

For each $a \in \operatorname{supp}(s)$, take some $f_a \in \partial c(\mu_a^s)$. Shifting f_a by a constant if necessary—which clearly preserves $f_a \in \partial c(\mu_a^s)$—assume without loss that $\int f_a(\omega)\ \mu_a^s(\mathrm{d}\omega) = c(\mu_a^s)$. Let us see $u \in \mathcal{U}$ given by

$$u_a = \begin{cases} f_a & : \ a \in \operatorname{supp}(s) \\ \min c(\Delta\Omega)\mathbf{1} & : \ a \in A \setminus \operatorname{supp}(s) \end{cases}$$

yields the desired payoff ranking. Indeed, for any $s' \in S$, Lemma 13 tells us

$$\begin{aligned} d_s^+ \kappa_c(s') &\geq \sum_{a \in \operatorname{supp}(s)} \mathbb{E}\left[(s'_a - s_a) u_a\right] + \sum_{a \in \operatorname{supp}(s') \setminus \operatorname{supp}(s)} \mathbb{E}\left[(s'_a - s_a)\ c(\mu_a^{s'})\right] \\ &\geq \mathbb{E}\left[(s' - s) \cdot u\right], \end{aligned}$$

as required. □

B.3. Convexity and Monotonicity

The following lemma constructs a "convex monotone envelope" of an information cost function $\hat{C}$ and establishes some regularity properties of the same.

Lemma 20. *Suppose $\hat{C} : \mathcal{P} \to \mathbb{R} \cup \{\infty\}$ is proper and lower semicontinuous. Then,*

$$\begin{aligned} C : \mathcal{P} &\to \mathbb{R} \cup \{\infty\} \\ p &\mapsto \min_{Q \in \Delta\mathcal{P}:\ \int q\, Q(\mathrm{d}q) \succeq p} \int \hat{C}(q)\, Q(\mathrm{d}q) \end{aligned}$$

is well defined (i.e., the defining minimum exists), proper, lower semicontinuous, convex, and monotone.

Proof. Given the HLPBSSC theorem, the relation $\succeq$ is continuous (i.e., a closed subset of $\mathcal{P}^2$). We can therefore apply a version of the maximum theorem (e.g., Lemma 17.30 from Aliprantis and Border, 2006) because the barycenter map is continuous and $\hat{C}$ is lower semicontinuous: the map $C : \mathcal{P} \to \mathbb{R} \cup \{\infty\}$ is well defined and lower semicontinuous.[35] Moreover, C is monotone because $\succeq$ is transitive, and it is proper because $\hat{C}$ is proper and $C \leq \hat{C}$.

Finally, we turn to convexity. The HLPBSSC theorem implies $\succeq$ is a convex subset of $\mathcal{P}^2$, and so the correspondence $\mathcal{P} \rightrightarrows \Delta\mathcal{P}$ given by $p \mapsto \{Q \in \Delta\mathcal{P} :\ \int q\, Q(\mathrm{d}q) \succeq p\}$ has a convex graph. Hence, C is a (weakly) convex function. □

Proof of Proposition 8. Let C be the function defined in the statement of Lemma 20, which (by that lemma) satisfies all of our standing assumptions on information cost functions.

Take any SCR s. Lemma 1 implies a given $Q \in \Delta\mathcal{P}$ can induce s if and only if $\int p\, Q(\mathrm{d}p) \succeq p^s$. Hence, by Lemma 20, some $Q \in \Delta\mathcal{P}$ of minimum average cost $\int \hat{C}(p)\, Q(\mathrm{d}p)$ can induce s, and this cost is exactly equal to $C(p^s)$. The result follows. □

B.4. Continuous Choice with Bounded Utilities

We begin with an abstract result on $\mathcal{V}$-rationalizability for the case in which $\mathcal{V}$ is a well-behaved linear subspace of $\mathcal{U}$.

[35] The map takes values in $\mathbb{R} \cup \{\infty\}$ rather than $\mathbb{R}$, but the cited lemma can be applied because $\mathbb{R} \cup \{\infty\}$ is homeomorphic to a subset of $\mathbb{R}$ via a strictly increasing transformation.

Lemma 21. *Suppose $\mathcal{V} \subseteq \mathcal{U}$ is a linear subspace, $(\tilde{\mathcal{S}}, \|\cdot\|)$ is a normed space with $\tilde{\mathcal{S}} \supseteq \mathcal{S}$, and $\{\varphi|_S : \varphi : \tilde{\mathcal{S}} \to \mathbb{R}$ linear continuous$\}$ is the set of maps $S \to \mathbb{R}$ given by $s \mapsto \mathbb{E}[u \cdot s]$ for $u \in \mathcal{V}$. Then, an SCR s is $\mathcal{V}$-rationalizable if and only if*

$$\inf_{s' \in S^\kappa:\ s' \neq s} \frac{\kappa(s') - \kappa(s)}{\|s' - s\|} > -\infty.$$

Proof. Fix $s \in S$. Viewing κ as a function on $\tilde{\mathcal{S}}$ (by letting it take value ∞ on $\tilde{\mathcal{S}} \setminus S$), the subdifferential of κ at s takes the form

$$\partial_{\tilde{\mathcal{S}}} \kappa(s) = \left\{ \varphi \in \tilde{\mathcal{S}}^* :\ \kappa(\tilde{s}) \geq \kappa(s) + \varphi(\tilde{s} - s)\ \forall \tilde{s} \in \tilde{\mathcal{S}} \right\}.$$

By hypothesis, $\{\varphi|_S :\ \varphi \in \tilde{\mathcal{S}}^*\}$ is the set of functions $S \to \mathbb{R}$ that are given by $s \mapsto \mathbb{E}[u \cdot s]$ for some $u \in \mathcal{V}$. Hence,

$$\begin{aligned}
& \partial_{\tilde{\mathcal{S}}} \kappa(s) \neq \varnothing \\
\iff \quad & \{\varphi|_S :\ \varphi \in \partial_{\tilde{\mathcal{S}}} \kappa(s)\} \neq \varnothing \\
\iff \quad & \exists u \in \mathcal{V} \text{ such that } \forall s' \in S,\ \kappa(s') \geq \kappa(s) + \mathbb{E}[u \cdot (s' - s)] \\
\iff \quad & s \text{ is } \mathcal{V}\text{-rationalizable.}
\end{aligned}$$

Therefore, the lemma follows immediately from Gale's (1967) duality theorem. □

Proof of Proposition 9. Letting $\tilde{\mathcal{S}} = L^1(\mu_0)^A$, whose dual space is naturally identified with $\mathcal{V} = L^\infty(\mu_0)^A$, the result follows directly from the first part of Lemma 21. □

Although we have applied Lemma 21 to determine which SCRs are rationalized by some bounded utility, one can vary the norm (generating distinct bounded-steepness conditions) to characterize rationalizability with respect to different classes of utilities. For example, κ exhibits bounded steepness with respect to the $L^2(\mu_0)$ norm at exactly the SCRs that can be rationalized by a finite-variance utility. Similarly, bounded steepness of κ with respect to the Kantorovich-Rubinstein norm characterizes rationalizability by a Lipschitz utility.

In contrast to the previous examples, Lemma 21 cannot be applied directly to the case in which $\mathcal{V}$ is the space of continuous functions $\Omega \to \mathbb{R}$. The reason is that, whereas the above examples relied on viewing $\mathcal{S}$ as a subset of some $\tilde{\mathcal{S}}$ whose dual was naturally identified with $\mathcal{V}$, the space of continuous functions is typically not a dual. Indeed, if Ω is infinite with finitely many connected components (e.g., if it is $[0, 1]$), one can easily show (via the

Banach-Alaoglu theorem) the space of continuous functions is not the dual of any normed space.

B.5. Costly Stochastic Choice

Let us state the analogue of Theorem 3 that we reported in section 7.

Proposition 11. *Suppose $\tilde{\kappa}$ is strictly convex on its domain $S^{\tilde{\kappa}}$ and is finite and differentiable at every conditionally full-support SCR. Further, suppose every full-support $s \in S^{\tilde{\kappa}}$ that does not have conditionally full support admits some $s' \in S$ such that $d_s^+\tilde{\kappa}(s') = -\infty$. Then, SCRs yielding unique subset predictions are weak* dense in $S^{\tilde{\kappa}}$.*

We now build up to the proof of the above proposition.

Remark 2. *In the analysis of this section, we apply Lemmas 3 and 4 and Propositions 1 and 2 to $\tilde{\kappa}$. All of these results were proven under the hypothesis that κ is proper, convex, and weak* lower semicontinuous (established in Corollary 1 for κ and directly assumed for $\tilde{\kappa}$). In particular, inspection of the proofs of these results shows they do not use the fact that κ is derived from C, and so the results can be applied to $\tilde{\kappa}$ without change.*

The following lemma adapts Lemma 16 to the simpler setting in which $\tilde{\kappa}$ is assumed differentiable.

Lemma 22. *Let $s \in S^{\tilde{\kappa}}$ have $\operatorname{supp}(s) = A$, and suppose u is a derivative of $\tilde{\kappa}$ at s. The following are equivalent for $v \in \mathcal{U}$:*

(i) SCR s is v-rationalizable; that is, $s \in \operatorname{argmax}_{t \in S} [\mathbb{E}[v \cdot t] - \tilde{\kappa}(t)]$.

(ii) Some $\lambda \in L^1(\mu_0)$ and $\gamma \in L^1(\mu_0)_+^A$ exist such that every $a \in A$ has

$$\begin{aligned} v_a &= \lambda - \gamma_a + u_a, \\ s_a \gamma_a &= 0. \end{aligned}$$

Proof. Let $u^s := v - u \in \mathcal{U}$. Lemma 4 tells us s is v-rationalizable if and only if every $s' \in S^\kappa$ has $d_s^+\tilde{\kappa}(s') \geq \mathbb{E}[v \cdot (s' - s)]$. But, that u is a derivative of $\tilde{\kappa}$ at s implies $d_s^+\tilde{\kappa}(s') = \mathbb{E}[u \cdot (s' - s)]$. We can therefore write the optimality condition as $s \in \operatorname{argmax}_{s' \in S} \mathbb{E}[u^s \cdot s']$. As in the proof of Lemma 16, this condition is equivalent to requiring some $\lambda \in L^1(\mu_0)$ and $\gamma \in L^1(\mu_0)_+^A$ to have $u^s = (\lambda - \gamma_a)_{a \in A}$ and $(\gamma_a s_a)_{a \in A} = 0$. $\square$

The following analogue of Corollary 3 is an immediate consequence of Lemma 22.

Corollary 4. *Let the SCR s have conditionally full support. If $\tilde{\kappa}$ is differentiable at s, and u and u' both rationalize s, some $\lambda \in L^1(\mu_0)$ exists such that*

$$u_a = u'_a + \lambda \; \forall a \in A.$$

Next, we record an analogue of Lemma 17.

Lemma 23. *If $s \in S^{\tilde{\kappa}}$ admits a $s' \in S$ such that $d_s^+ \tilde{\kappa}(s') = -\infty$, then s is not rationalizable.*

Proof. For every $\epsilon \in (0,1)$, let $t^\epsilon = s + \epsilon(s' - s)$. Every $u \in \mathcal{U}$ then has

$$\frac{1}{\epsilon}\left\{\mathbb{E}\left[u \cdot t^\epsilon\right] - \tilde{\kappa}(t^\epsilon) - \left[\mathbb{E}\left[u \cdot s\right] - \tilde{\kappa}(s)\right]\right\} = \mathbb{E}\left[u \cdot (s' - s)\right] - \frac{1}{\epsilon}\left[\tilde{\kappa}(t^\epsilon) - \tilde{\kappa}(s)\right] \xrightarrow{\epsilon \searrow 0} -\infty.$$

Thus, for all sufficiently small $\epsilon \in (0,1)$, the agent's objective must be strictly higher under t^ϵ than under s. □

Next, we provide an analogue of Lemma 18.

Lemma 24. *Suppose $\tilde{\kappa}$ is finite and differentiable at every conditionally full-support SCR, every full-support $s \in S^{\tilde{\kappa}}$ that does not have conditionally full support admits some $s' \in S$ such that $d_s^+ \tilde{\kappa}(s') = -\infty$, and some norm-dense subset S_1 of $S^{\tilde{\kappa}}$ exists that comprises only uniquely rationalizable SCRs. Then, the set of SCRs yielding unique subset predictions is weak* dense in S.*

Proof. The proof of Lemma 18 (which invokes Lemmas 3 and 4 and Propositions 1 and 2) can be applied nearly verbatim, with three minor differences. First, the step invoking Assumption A2(i) is replaced with the hypothesis that the set $S^{\tilde{\kappa}}$ contains all conditionally full-support SCRs, because the latter set is dense in S. Second, we invoke Lemma 23 instead of Lemma 17. Third, we invoke Corollary 4 instead of Corollary 3. □

Finally, we can prove the main result of this section.

Proof of Proposition 11. By Proposition 2 and strict convexity of $\tilde{\kappa}$, the set S_1 of uniquely rationalizable SCRs is norm dense in $S^{\tilde{\kappa}}$. The proposition then follows from Lemma 24. □

C. Online Appendix: Auxiliary Material

C.1. Omitted Proofs for Auxiliary Results

This section provides proofs of some additional results that were stated in the previous appendices without proof.

Proof of Corollary 2. Say $A = \{0, 1\}$ without loss, and note $s \in S$ generates state-independent behavior if and only if $p^s = \delta_{\mu_0}$. We therefore require that if $s \in S$ is optimal and $p^s \neq \delta_{\mu_0}$, then s is uniquely optimal. But $p^s \neq \delta_{\mu_0}$ for $p^s \in \mathcal{P}$ implies both that $p_0^s, p_1^s > 0$ and that μ_0^s, μ_1^s are distinct, and hence affinely independent. Proposition 3 therefore applies directly.

Letting $p^* := p^{s^*}$ for some u-rationalizable SCR s^*, the above argument tells us $p^s = p^*$ for every u-rationalizable SCR s. Therefore (in light of Lemma 1), C being strictly monotone means every optimal strategy entails information policy exactly p^*, as required. $\square$

Proof of Fact 1. Toward establishing convexity of c, fix any $\mu_1, \mu_2 \in \Delta\Omega$ and $\beta \in (0, 1)$, and let $\mu = (1-\beta)\mu_1 + \beta\mu_2$. Let $(\mu_1^n, \mu_2^n)_{n\in\mathbb{N}}$ be a sequence of pairs of simply drawn posteriors converging to (μ_1, μ_2), which exists by Lipnowski and Mathevet's (2018) Lemma 2. By definition of a simply drawn posterior, some sequence $(\zeta_1^n, \zeta_2^n)_{n\in\mathbb{N}}$ of pairs from $\mathbb{R}_{++}$ exists such that $\zeta_i^n \mu_i^n \leq \mu_0$ for all $i \in \{1, 2\}$ and $n \in \mathbb{N}$.

Now, consider any $n \in \mathbb{N}$. Take $\zeta^n := \frac{1}{2}\min\{\zeta_1^n, \zeta_2^n\}$ and $\mu^n := (1-\beta)\mu_1^n + \beta\mu_2^n$ for $n \in \mathbb{N}$. Observe $2\zeta^n\mu^n \leq (1-\beta)\zeta_1^n\mu_1^n + \beta\zeta_2^n\mu_2^n \leq \mu_0$, and so $\hat{\mu}^n = \frac{\mu_0 - \zeta^n\mu^n}{1-\zeta^n}$ is a well-defined simply drawn posterior. Therefore, both

$$p_n := \zeta^n\delta_{\mu^n} + (1-\zeta^n)\delta_{\hat{\mu}^n} \text{ and } \tilde{p}_n := \zeta^n\left[(1-\beta)\delta_{\mu_1^n} + \beta\delta_{\mu_2^n}\right] + (1-\zeta^n)\delta_{\hat{\mu}^n}$$

are in $\mathcal{P}^F$. Because $p \in \text{feas}\, C$, it follows that $p + \epsilon(p_n - p)$ and $p + \epsilon(\tilde{p}_n - p)$ are both in $\mathcal{P}^C$ for all sufficiently small $\epsilon > 0$. Hence, because $\tilde{p}_n \succeq p_n$ by construction, we have (because C is monotone)

$$\begin{aligned}
0 &\leq \liminf_{\epsilon \searrow 0} \frac{1}{\epsilon}\left[C(p + \epsilon(\tilde{p}_n - p)) - C(p + \epsilon(p_n - p))\right] \\
&= \liminf_{\epsilon \searrow 0} \frac{1}{\epsilon}\left\{\left[C(p + \epsilon(\tilde{p}_n - p)) - C(p)\right] - \left[C(p + \epsilon(p_n - p)) - C(p)\right]\right\} \\
&= \int c(\mu)\,(\tilde{p}_n - p)(\mathrm{d}\mu) - \int c(\mu)\,(p_n - p)(\mathrm{d}\mu) \\
&= \zeta^n\left[(1-\beta)c(\mu_1^n) + \beta c(\mu_2^n) - c(\mu^n)\right].
\end{aligned}$$

Finally, that $(1-\beta)c(\mu_1^n)+\beta c(\mu_2^n)\geq c(\mu^n)$ for every $n\in\mathbb{N}$ yields $(1-\beta)c(\mu_1)+\beta c(\mu_2)\geq c(\mu)$ because c is continuous. The lemma follows. $\square$

Proof of Fact 2. Let $f\in L^1(\mu_0)$ be the difference of two derivatives of $c\in\mathcal{C}$ at the simply drawn μ.[36] By hypothesis, every simple $\tilde{\mu}\in\Delta\Omega$ has $\int f(\omega)\,(\tilde{\mu}-\mu)(\mathrm{d}\omega)=0$, and hence, $\int f(\omega)\,\tilde{\mu}(\mathrm{d}\omega)=\bar{f}:=\int f(\omega)\,\mu(\mathrm{d}\omega)$.

Because $\int f(\omega)\,\mu_0(\mathrm{d}\omega)=\bar{f}$, the event $\hat{\Omega}:=\{f\geq\bar{f}\}\subseteq\Omega$ has $\mu_0(\hat{\Omega})>0$. Let $\tilde{\mu}:=(\mu_0|\hat{\Omega})$ denote the conditional measure, which is a simply drawn posterior. Therefore,

$$0=\mu_0(\hat{\Omega})\int\left[f(\omega)-\bar{f}\right]\,\tilde{\mu}(\mathrm{d}\omega)=\mathbb{E}\left[\mathbf{1}_{\hat{\Omega}}(f-\bar{f})\right].$$

An expectation of a nonnegative random variable can be zero only if the random variable is almost surely zero, yielding two consequences. First, $\mathbf{1}_{\hat{\Omega}}(f-\bar{f})=0\in L^1(\mu_0)$, so that $f\leq\bar{f}$. Second, that $f\leq\bar{f}$ and $\mathbb{E}[f]=\bar{f}$ implies $f=\bar{f}\in L^1(\mu_0)$, a constant.

Finally, that every derivative of c at μ has μ-expectation $c(\mu)$ then implies $\bar{f}=0$.[37] $\square$

Proof of Fact 3. Every $q\in\mathcal{P}$ has $q\succeq\delta_{\mu_0}$, and so $p+\epsilon(q-p)\succeq p+\epsilon(\delta_{\mu_0}-p)$ for any $\epsilon\in(0,1)$. Hence,

$$\begin{aligned}
&\inf_{q\in\mathcal{P}^C,\ \epsilon\in(0,1)} && \frac{1}{\epsilon}\left[C(p+\epsilon(q-p))-C(p)\right]\\
=\ &\inf_{\epsilon\in(0,1)} && \frac{1}{\epsilon}\left[C(p+\epsilon(\delta_{\mu_0}-p))-C(p)\right]\ \text{(by monotonicity)}\\
=\ &\lim_{\epsilon\searrow 0} && \frac{1}{\epsilon}\left[C(p+\epsilon(\delta_{\mu_0}-p))-C(p)\right]\ \text{(by convexity)},
\end{aligned}$$

which is equal to $d_p^+C(\delta_{\mu_0})$ by definition. $\square$

C.2. On Inducible Belief Distributions

In this section, we validate footnote 4's claim that belief distributions and signal structures are equivalent formalisms. To state the equivalence, let us invest in some terminology.

Definition 1. *Given a Polish space M, a **signal** is a measurable map $\tau:\Omega\to\Delta M$, and a **belief map** is a measurable map $\pi:M\to\Delta\Omega$. For a given pair (τ,π) of such maps,*

[36] Note that the proof relies only on the "Newton quotient" property of ∇c_μ, together with its average value normalization, and so does not require that $c\in\mathcal{C}$.

[37] If we did not require the normalization that $\int\nabla c_\mu(\omega)\,\mu(\mathrm{d}\omega)=c(\mu)$—which would not affect the definition of differentiability—the derivative would be unique only up to addition of a constant function.

- *Say the pair (τ, π) is **Bayes consistent** if every $\hat{\Omega} \subseteq \Omega$ and $\hat{M} \subseteq M$ have*

$$\int_{\hat{\Omega}} \tau(\hat{M}|\omega)\, \mu_0(\mathrm{d}\omega) = \int \int_{\hat{M}} \pi(\hat{\Omega}|m)\, \tau(\mathrm{d}m|\omega)\, \mu_0(\mathrm{d}\omega).$$

- *Say the pair (τ, π) **generates** $p \in \Delta\Delta\Omega$ if every $B \subseteq \Delta\Omega$ has*

$$\int \tau\left(\pi^{-1}(B) \mid \omega\right)\, \mu_0(\mathrm{d}\omega) = p(B).$$

Note every signal admits some Bayes-consistent belief map, and that updated beliefs are almost surely unique, and hence unique in distribution. These observations amount to recording a standard disintegration result in present notation.

Fact 4. *Suppose M is Polish and that τ is a signal. Define $\mathbb{P} \in \Delta(\Omega \times M)$ by letting*

$$\mathbb{P}(\hat{\Omega} \times \hat{M}) := \int_{\hat{\Omega}} \tau(\hat{M}|\omega)\, \mu_0(\mathrm{d}\omega)$$

for every measurable $\hat{\Omega} \subseteq \Omega$ and $\hat{M} \subseteq M$. Define $\mathbb{P}_M := \mathrm{marg}_M \mathbb{P}$.

(i) Some belief map π exists such that the pair (τ, π) is Bayes consistent.

(ii) If belief maps $\pi, \tilde{\pi}$ are such that both (τ, π) and $(\tau, \tilde{\pi})$ are Bayes consistent, $\pi(\hat{\Omega}|m) = \tilde{\pi}(\hat{\Omega}|m)$ for $\mathbb{P}_M$-almost every m.

(iii) Suppose belief maps $\pi, \tilde{\pi}$ are such both (τ, π) and $(\tau, \tilde{\pi})$ are Bayes consistent; and suppose $p, \tilde{p} \in \Delta\Delta\Omega$ are such that (τ, π) and $(\tau, \tilde{\pi})$ generate p and $\tilde{p}$, respectively. Then, $p = \tilde{p}$.

Proof. Given a belief map π, note the pair (τ, π) is Bayes consistent if and only if $\mathbb{P}(\hat{\Omega} \times \hat{M}) = \int_{\hat{M}} \pi(\hat{\Omega}|m)\, \mathbb{P}_M(\mathrm{d}m)$ for every measurable $\hat{\Omega} \subseteq \Omega$ and $\hat{M} \subseteq M$. Hence, (i) and (ii) both follow directly from the disintegration theorem (Kallenberg, 2017, Theorem 1.23).[38] Finally, any two ($\Delta\Omega$-valued) random variables on a probability space that agree almost surely must have the same distribution; hence, (iii) follows directly from (ii). □

The following result shows every signal generates (when pairing it with belief updating to make it Bayes consistent) a unique belief distribution, and that a belief distribution can be generated in this way if and only if it averages to the prior. This classic result—whose

[38] By (Aliprantis and Border, 2006, Theorem 19.7), probability kernels as in Kallenberg (2017) coincide with measurable maps into the space of measures.

proof we include for the sake of completeness—is again a straightforward consequence of the disintegration theorem.

Fact 5. *Let M be uncountable and Polish, and let $p \in \Delta\Delta\Omega$. The following are equivalent:*[39]

(i) Some Bayes-consistent pair (τ, π) generates p.

(ii) Some signal τ is such that every Bayes-consistent (τ, π) generates p.

(iii) The distribution p averages to μ_0, that is, $p \in \mathcal{P}$.[40]

Proof. Given any signal τ, Fact 4 tells us some Bayes-consistent pair includes τ and that no two such pairs generate distinct belief distributions. Hence, (i) is equivalent to (ii). In what follows, we show (i) is equivalent to (iii).

To see (i) implies (iii), suppose (τ, π) is Bayes consistent and generates p. Toward (iii), consider any measurable $\hat{\Omega} \subseteq \Omega$. Applying Bayes consistency with $\hat{M} = M$ yields $\mu_0(\hat{\Omega}) = \int\int \pi(\hat{\Omega}|m)\ \tau(\mathrm{d}m|\omega)\ \mu_0(\mathrm{d}\omega)$; and because (τ, π) generates p, any bounded integrable $f : \Delta\Omega \to \mathbb{R}$ has $\int\int f\left(\pi(m)\right)\ \tau(\mathrm{d}m|\omega)\ \mu_0(\mathrm{d}\omega) = \int f(\mu)\ p(\mathrm{d}\mu)$. Hence, using $f(\mu) := \mu(\hat{\Omega})$ gives $\mu_0(\hat{\Omega}) = \int \mu(\hat{\Omega})\ p(\mathrm{d}\mu)$, delivering (iii).

Conversely, suppose (iii) holds, from which we establish (i). By the Borel isomorphism theorem (Srivastava, 2008, Theorem 3.3.13), some bimeasurable surjection $\pi : M \to \Delta\Omega$ exists.[41] Now, define the measure $\mathbb{P} \in \Delta(\Omega \times M)$ by letting $\mathbb{P}(\hat{\Omega} \times \hat{M}) := \int_{\pi(\hat{M})} \mu(\hat{\Omega})\ p(\mathrm{d}\mu)$ for every measurable $\hat{\Omega} \subseteq \Omega$ and $\hat{M} \subseteq M$. By (iii) and because π is surjective, the marginal of $\mathbb{P}$ on its first coordinate is μ_0. Hence, the disintegration theorem (Kallenberg, 2017, Theorem 1.23) delivers some measurable $\tau : \Omega \to \Delta M$ such that $\mathbb{P}(\hat{\Omega} \times \hat{M}) = \int_{\hat{\Omega}} \tau(\hat{M}|\omega)\ \mu_0(\mathrm{d}\omega)$ for every measurable $\hat{\Omega} \subseteq \Omega$ and $\hat{M} \subseteq M$. Let us see that (τ, π) witnesses (i). Indeed, for any measurable $\hat{\Omega} \subseteq \Omega$ and $\hat{M} \subseteq M$, combining the disintegration property defining τ with the definition of $\mathbb{P}$ implies

$$\int_{\hat{\Omega}} \tau(\hat{M}|\omega)\ \mu_0(\mathrm{d}\omega) = \int_{\pi(\hat{M})} \mu(\hat{\Omega})\ p(\mathrm{d}\mu). \tag{6}$$

Specializing equation (6) to the case of $\hat{\Omega} = \Omega$ and $\hat{M} = \pi^{-1}(B)$ for some measurable $B \subseteq \Delta\Omega$ tells us, because π is surjective, that (τ, π) generates p. All that remains now is

[39] As the proof demonstrates, (i) and (ii) are equivalent and imply (iii), even if M is not assumed uncountable. Moreover, the proof (along with the proof of Fact 4) applies verbatim to Polish, non-compact Ω.

[40] Recall elements of $\mathcal{P}$ are those $p \in \Delta\Delta\Omega$ with barycenter μ_0 or, equivalently, with $\int \mu(\hat{\Omega})\ p(\mathrm{d}\mu) = \mu_0(\hat{\Omega})$ for every measurable $\hat{\Omega} \subseteq \Omega$.

[41] Moreover, in the special case in which $M \supseteq \Delta\Omega$, we can take π with $\pi(\mu) = \mu$ for each $\mu \in \Delta\Omega$, generating a witnessing (τ, π) in which each signal realization from $\Delta\Omega$ leads to itself as the updated belief.

to verify (τ, π) is Bayes consistent. To show it is, define the map $\mathbb{P}_m \in \Delta M$ by letting $\mathbb{P}_M \in \Delta M := \int \tau(\hat{M}|\omega)\, \mu_0(\mathrm{d}\omega)$ for every measurable $\hat{M} \subseteq M$. We know (τ, π) generates p; that is, $\mathbb{P}_M \circ \pi^{-1} = p$. Hence, every measurable $\hat{\Omega} \subseteq \Omega$ and $\hat{M} \subseteq M$ have

$$\int \int_{\hat{M}} \pi(\hat{\Omega}|m)\, \tau(\mathrm{d}m|\omega)\, \mu_0(\mathrm{d}\omega) = \int_{\hat{M}} \pi(\hat{\Omega}|m)\, \mathbb{P}_M(\mathrm{d}m) = \int_{\pi(\hat{M})} \mu(\hat{\Omega})\, p(\mathrm{d}\mu).$$

Bayes consistency then follows from equation (6). $\square$