Theoretical Economics 12 (2017), 621–654 1555-7561/20170621

Rationally inattentive preferences and hidden information costs

Henrique de Oliveira
Department of Economics, Penn State University

Tommaso Denti
Department of Economics, MIT

Maximilian Mihm
Division of Social Science, NYU Abu Dhabi

Kemal Ozbek
School of Economics and Finance, University of St. Andrews

We show how information acquisition costs can be identified using observable choice data. Identifying information costs from behavior is especially relevant when these costs depend on factors—such as time, effort, and cognitive resources—that are difficult to observe directly, as in models of rational inattention. Using willingness-to-pay data for opportunity sets—which require more or less information to make choices—we establish a set of canonical properties that are necessary and sufficient to identify information costs. We also provide an axiomatic characterization of the induced rationally inattentive preferences and show how they reveal the amount of information a decision-maker acquires.

Keywords. Information costs, Blackwell order, information acquisition, menu choice, rational inattention.

JEL classification. D81, D83.

Henrique de Oliveira: henrique@psu.edu
Tommaso Denti: tdenti@princeton.edu
Maximilian Mihm: max.mihm@nyu.edu
Kemal Ozbek: kemal.ozbek@st-andrews.ac.uk

This paper merges our independent works, previously circulated as "Axiomatic Foundations for Rational Inattention" (de Oliveira), "Rational Inattention: A Decision-Theoretic Perspective" (Denti), and "Decision Making with Rational Inattention" (Mihm and Ozbek). We are grateful to Larry Blume, Simone Cerreia-Vioglio, Eddie Dekel, Fabio Maccheroni, Paola Manzini, Massimo Marinacci, Marco Mariotti, Tapan Mitra, Todd Sarver, and Joerg Stoye for invaluable guidance and support. We also thank Nabil Al-Najjar, Andrew Caplin, Aaron Bodoh-Creed, Giulia Brancaccio, Enrico Cantoni, Mark Dean, David Easley, Ludovica Gazzé, Peter Klibanoff, John Leahy, Samreen Malik, Jawwad Noor, Efe Ok, Pietro Ortoleva, Luciano Pomatto, Debraj Ray, Marciano Siniscalchi, Juuso Toikka, Mirko Wiederholt, Muhamet Yildiz, various seminar and conference participants, the editor and three anonymous reviewers for helpful comments and suggestions.

Copyright © 2017 The Authors. Theoretical Economics. The Econometric Society. Licensed under the Creative Commons Attribution-NonCommercial License 3.0. Available at http://econtheory.org.
DOI: 10.3982/TE2302

1. Introduction

In many economic settings, decisions are not made on the basis of some fixed prior information: what to know is itself a choice variable. While information improves decision-making, acquiring information is also costly, and agents therefore have to balance the benefits and costs.

A challenge in the study of information acquisition is how to identify the information costs an agent faces. These costs determine the amount and type of information the agent acquires, and are therefore crucial for comparative statics, welfare, and policy analysis. However, the costs can depend on many factors that are difficult to measure directly. A prominent example is the recent literature on rational inattention (Sims 1998, 2003), where agents are assumed to have access to an abundance of information and choose what information to pay attention to. In other words, they solve an information-acquisition problem where the costs represent factors—such as time, effort, and cognitive resources—that are difficult to observe.[1]

In this paper, we consider a general model of information acquisition and show how hidden information costs can be identified from *menu-choice* data. We take the perspective of an agent when she chooses a menu—an opportunity set—"today," with the prospect of acquiring information "tomorrow" before selecting an alternative from the menu. Preferences over menus reveal the trade-off the agent faces when she balances the benefits of information (which depend on the menu) against the costs of information (which are not directly observable). We call the resulting preference relation over menus a *rationally inattentive preference*.

We establish that three canonical properties are enough to uniquely identify information costs: (i) no information is costless, (ii) Blackwell monotonicity, and (iii) convexity. These are standard properties of information costs used in many applications, such as the rational inattention literature. We show that the properties impose no observable restrictions on choice data and therefore constitute a natural benchmark for models of information acquisition with hidden information costs. Moreover, we provide an explicit formula showing how the canonical information costs can be constructed with willingness-to-pay data, which could be collected in dynamic choice environments or generated in experimental settings.

Figure 1 illustrates the timeline of our framework. Today, the agent chooses a menu being uncertain about the future state of fundamentals. Tomorrow, a state realizes and

Figure 1. Timeline.

[1] See, for example, Wiederholt (2010) for a review of the rational inattention literature, and Veldkamp (2011) for recent developments in the economics of information acquisition.

information becomes available. The agent is able to acquire information so as to reduce uncertainty about the state before choosing an alternative with state-contingent payoffs. The value of a menu therefore depends on the set of available choices and the agent's costs of information. Intuitively, an agent with high costs is willing to pay more for menus in which not as much information is needed to make a better choice. An agent with lower costs is willing to pay less, because she is better able to acquire the information needed to reduce uncertainty about the state.

This framework can be viewed as a "snapshot" of a dynamic choice environment, where agents make choices every period that affect both current well-being and future choice opportunities. For example, in the consumption–saving problems studied in the rational inattention literature (e.g., Sims 2003, Luo 2010, and Tutino 2013), an agent acquires information about a random income source in each period, and then decides how much to consume and how much to save. The saving decision in a particular period affects future consumption opportunities, and can therefore be viewed as the choice of a menu of future consumption–saving plans, which is made before information about future income is acquired. The choice data could also be generated in experimental settings. For example, Gabaix et al. (2006), Caplin and Dean (2014), and Cheremukhin et al. (2015) suggest experimental designs to study rational inattention in the laboratory, and our analysis shows how "willingness-to-pay" data collected in such experiments can be used to infer the hidden information costs of participants.

In addition to our identification result, we provide an axiomatic characterization of rationally inattentive preferences. The axioms can be used to verify when a menu-choice data set is consistent with a model of information acquisition, as depicted in Figure 1. Applying methods developed in Maccheroni et al. (2006), we show that rationally inattentive preferences can be viewed as a class of the costly contemplation preferences in Ergin and Saver (2010), and a generalization of the subjective-learning preferences in Dillenberger et al. (2014).

Finally, we use the class of canonical information costs for comparative statics, and show how information costs and information acquisition—which are both typically not observable—can be compared across agents with menu-choice data. Information costs can be compared following a standard approach, where agents rank menus that offer them flexibility tomorrow versus menus that force them to commit to an alternative today (as in Dekel et al. 2001, Ergin and Saver 2010, Dillenberger et al. 2014). However, such comparisons are not enough to reveal how much information agents acquire, an important question in applications of rational inattention. We therefore introduce a criterion to evaluate when one menu offers a higher "premium for information" than another (in the sense that additional information is more valuable), and show that an agent acquires more information whenever she has a stronger preference for menus with a higher information premium.

Our analysis complements testable implications of rational inattention developed in a number of recent papers (e.g., Caplin and Dean 2015, Ellis 2016, Matejka and McKay 2015), which look at an agent's choices after she acquires information. In particular, Caplin and Dean (2015) show that canonical information costs are also without loss of generality for state-dependent stochastic choice data, but whether these properties are

sufficient to identify a unique cost function with their data remains an open question. Our framework allows us to obtain a full identification of all model parameters, without exogenous restrictions on the set of available signals or additional assumptions about the agent's utility function or prior beliefs. Moreover, since ex post verification of rational inattention requires the stochastic choice data corresponding to every state of world, our analysis highlights a complementary choice data set that can be used to elicit preference parameters and predict ex post behavior when collecting state-dependent stochastic data is difficult or infeasible.

The paper is organized as follows. In Section 2, we present our menu-choice framework and relate it to a general model of information acquisition. Section 3 characterizes the testable implications of the model for menu choice. In Section 4, we present our identification results: we introduce the canonical properties of an information cost function, and show that they are necessary and sufficient to identify all model parameters. Section 5 shows how menu-choice data reveal how much information agents acquire. Section 6 concludes. Proofs are provided in the Appendix.

2. Preliminaries

In this section, we introduce our choice framework of menus of state-contingent alternatives. We then describe a general information-acquisition problem and define the induce preference relation over menus.

2.1 *Framework*

There is a finite set Ω of *states* and a set X of *outcomes*, consisting of simple (finite-support) lotteries on a set of deterministic prizes (such as money or consumption).[2] An (Anscombe–Aumann) *act* $f : \Omega \to X$ is a map from states into outcomes and the set of all acts is denoted $\mathcal{F}$. A *menu* $F \subset \mathcal{F}$ is a finite set of acts and $\mathbb{F}$ denotes the collection of all menus.

For $\alpha \in [0, 1]$, acts $f, g \in \mathcal{F}$, and menus $F, G \in \mathbb{F}$, we denote by $\alpha f + (1-\alpha)g$ the *mixed act* h such that

$$h(\omega) = \alpha f(\omega) + (1-\alpha)g(\omega) \quad \forall \omega \in \Omega,$$

and by $\alpha F + (1-\alpha)G$, we denote the *mixed menu* H such that

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

Our primitive is a binary relation $\succsim$ over the set of menus, which represents the preferences of a decision-maker (henceforth, DM). The asymmetric and symmetric parts of $\succsim$ are denoted $\succ$ and $\sim$, respectively. With some abuse of notation, we identify a singleton menu $\{f\}$ with the act $f \in \mathcal{F}$, and identify a constant act f such that $f(\omega) = x$ for all $\omega \in \Omega$ with the outcome $x \in X$.

[2] We assume that Ω is finite to simplify the exposition. Appendix A.3 provides the technical modifications required to accommodate a general measurable space.

To interpret the framework, we take the perspective of a DM when she chooses her menu (opportunity set), with the prospect of processing information before she selects an act (alternative). As a result, $F \succsim G$ should be interpreted as "confronted today with the choice between menus F and G, the DM (weakly) prefers F to G as her opportunity set for tomorrow."

The objective lotteries in X represent an additional source of uncertainty that realizes *after* the DM has chosen an act. Formally, the mixed menu $\alpha F + (1-\alpha)G$ is just another set of Anscombe–Aumman acts. However, as in Ergin and Saver (2010), it is useful to provide an interpretation in terms of *contingency plans*. To illustrate, suppose that the randomization α represents the toss of a coin (that lands on heads with probability α and tails with probability $(1-\alpha)$). When choosing an act from the menu $\alpha F + (1-\alpha)G$, the DM could make a contingency plan: choose $f \in F$ on heads and $g \in G$ on tails. Corresponding to this contingency plan, there is a mixed act $\alpha f + (1-\alpha)g \in \alpha F + (1-\alpha)G$ that, in each state ω, yields the outcome that f delivers when the coin lands on heads and the outcome that g delivers when the coin lands on tails. As such, $\alpha f + (1-\alpha)g$ can be interpreted as the contingency plan "choose f if heads and g if tails," and the menu $\alpha F + (1-\alpha)G$ can be interpreted as the set of all contingency plans that can be formed from menus F and G.

2.2 *The information-acquisition problem*

We consider a general information-acquisition problem, which describes how the DM chooses an act from a menu. Before a state of the world is realized, the DM has a prior $\bar{p} \in \Delta(\Omega)$ that represents her initial beliefs. After a state is realized, the DM can acquire a noisy signal that conveys additional information about the state. Each possible realization of the signal induces a posterior belief $p \in \Delta(\Omega)$ from the prior $\bar{p}$ via Bayes rule. Accordingly, a signal leads to a *distribution over posteriors* $\pi \in \Delta(\Delta(\Omega))$, which satisfies the Bayesian requirement that the expected posterior is equal to the prior. As a result, the collection of all possible signals is given by the set

$$\Pi(\bar{p}) = \left\{ \pi \in \Delta\big(\Delta(\Omega)\big) : \int_{\Delta(\Omega)} p\,\pi(dp) = \bar{p} \right\}.$$

The set of signals $\Pi(\bar{p})$ is partially ordered in terms of their "informativeness" by the well known ranking of Blackwell (1951, 1953), which in this context can be defined as follows.

Definition 1. Signal $\pi \in \Pi(\bar{p})$ is *Blackwell more informative* than signal $\rho \in \Pi(\bar{p})$, denoted $\pi \trianglerighteq \rho$, if

$$\int_{\Delta(\Omega)} \varphi(p)\,\pi(dp) \geq \int_{\Delta(\Omega)} \varphi(p)\,\rho(dp)$$

for every convex continuous function $\varphi : \Delta(\Omega) \to \mathbb{R}$.

Given a menu F, extracting a signal allows the DM to make a more informed choice from F because she can choose an act to maximize expected utility for each posterior

$p \in \Delta(\Omega)$. With a utility function $u : X \to \mathbb{R}$, the *benefit of information* for a signal $\pi \in \Pi(\bar{p})$ is, therefore,

$$b_F^u(\pi) = \int_{\Delta(\Omega)} \left[\max_{f \in F} \left(\int_\Omega u\big(f(\omega)\big)\, p(d\omega) \right) \right] \pi(dp).$$

Since the integrand in square brackets is a convex continuous function on $\Delta(\Omega)$, the benefits of information are increasing in the Blackwell order.

A rationally inattentive DM balances the benefit of information from a signal π against the cost for acquiring that signal. These costs are measured by an *information cost function* $c : \Pi(\bar{p}) \to [0, \infty]$, which associates a cost $c(\pi)$ to each signal $\pi \in \Pi(\bar{p})$. In the information-acquisition problem, the DM therefore chooses a signal π that maximizes the difference between benefits and costs of information $(b_F^u(\pi) - c(\pi))$.

2.3 *Rationally inattentive preferences*

In our framework, the DM chooses a menu today with the prospect of acquiring information tomorrow before she selects an act. We model information acquisition as illustrated above and study the induced preference relation over menus. A leading interpretation is that the DM is "rationally inattentive": she has access to an abundance of information and chooses what to pay attention to. We therefore call the induced preference relation over menus a *rationally inattentive preference*.

Definition 2. A binary relation $\succsim$ over menus is a rationally inattentive preference if it is represented by a functional $V : \mathbb{F} \to \mathbb{R}$, defined by

$$V(F) = \max_{\pi \in \Pi(\bar{p})} \big[b_F^u(\pi) - c(\pi) \big], \tag{1}$$

where $u : X \to \mathbb{R}$ is an unbounded affine utility function, $\bar{p} \in \Delta(\Omega)$ is a prior, and $c : \Pi(\bar{p}) \to [0, \infty]$ is a proper lower-semicontinuous information cost function.[3]

The assumptions on parameters $(u, \bar{p}, c)$ are standard. Properness (i.e., $c(\pi) < \infty$ for some π) and lower semicontinuity of c are the minimal assumptions required to ensure that the maximization over costly signals is well defined. The affinity of u corresponds to the assumption of von Neumann–Morgenstern utility over lotteries. Finally, the unboundedness of u ensures that the benefit of information is not bounded, which is important for our identification approach.

The following examples illustrate some special cases of rationally inattentive preferences, which are relevant in applications.

Constrained information In a constrained-information problem, the DM does not incur a cost of information but is limited to choose a signal from some nonempty compact set $\Gamma \subset \Pi(\bar{p})$. A *constrained-information preference* is therefore represented by $(u, \bar{p}, \Gamma)$, such that a menu F is evaluated by $V(F) = \max_{\pi \in \Gamma} b_F^u(\pi)$. Such preferences are a special case of Definition 2 where $c(\pi) = 0$ if $\pi \in \Gamma$ and $c(\pi) = \infty$ otherwise.

[3]The set $\Delta(\Delta(\Omega))$ is endowed with the weak* topology.

Subjective learning In a model of subjective learning, the DM acquires an exogenous signal $\pi^* \in \Pi(\bar{p})$ and cannot adjust her information depending on the choice problem. A *subjective-learning preference* (Dillenberger et al. 2014) is therefore represented by $(u, \bar{p}, \pi^*)$, such that a menu F is evaluated by $V(F) = b_F^u(\pi^*)$. Such preferences are a special case of Definition 2 where $c(\pi) = 0$ if $\pi = \pi^*$ and $c(\pi) = \infty$ otherwise.

Mutual information Sims (1998, 2003) suggests parametrizations of information costs using Shannon's *mutual information* (Cover and Thomas 2006, Chapter 2), which measures information in terms of expected entropy reduction:

$$I(\pi) = \int_{\Delta(\Omega)} \left(\int_{\Omega} \log\big[p(\omega)/\bar{p}(\omega)\big] p(d\omega) \right) \pi(dp).$$

For example, a linear specification, $c_\lambda(\pi) = \lambda I(\pi)$, represents costs in terms of the parameter $\lambda \geq 0$ that measures the unit costs of information, and a constraint specification

$$c_\kappa(\pi) = \begin{cases} 0 & \text{if } I(\pi) \leq \kappa, \\ \infty & \text{otherwise} \end{cases}$$

represents costs in terms of the parameter $\kappa \geq 0$ that measures a capacity constraint on information processing.

3. Characterization

In this section, we show that rationally inattentive preferences can be characterized by a simple set of axioms from the menu-choice literature. We first present the axioms, then discuss the representation theorem, and finally characterize some special cases.

3.1 *Axioms*

In the sequel, we consider a number of axioms from the menu-choice literature. The first three axioms are standard.

Axiom 1 (Weak order). *For all menus F, G, and H, (i) $F \succsim G$ or $G \succsim F$ and (ii) if $F \succsim G$ and $G \succsim H$, then $F \succsim H$.*

Axiom 2 (Continuity). *For all menus F, G, and H, the following sets are closed:*

$$\big\{\alpha \in [0,1] : \alpha F + (1-\alpha)G \succsim H\big\} \quad \text{and} \quad \big\{\alpha \in [0,1] : H \succsim \alpha F + (1-\alpha)G\big\}.$$

Axiom 3 (Unboundedness). *There are outcomes x and y, with $x \succ y$, such that for all $\alpha \in (0,1)$, there is an outcome z satisfying either $y \succ \alpha z + (1-\alpha)x$ or $\alpha z + (1-\alpha)y \succ x$.*

Axioms 1 and 2 ensure that preferences are complete, transitive, and continuous. Axiom 3 implies that preferences over outcomes are unbounded (see, e.g., Maccheroni et al. 2006, Lemma 29). The remaining axioms reflect distinctive features of the information-acquisition problem that defines a rationally inattentive preference.

First, in the information-acquisition problem, the DM chooses an optimal act conditional on the information conveyed by signal realizations. As a result, when choosing a menu, the DM exhibits a *preference for flexibility* (Kreps 1979): adding an act to a menu cannot make the DM worse off (since she can always ignore the act if it is not optimal).[4]

Axiom 4 (Preference for flexibility). *For all menus F and G, if $G \subset F$, then $F \succsim G$.*

Second, in the information acquisition problem, the DM chooses a signal to balance the benefits and costs of information for the specific menu. In our framework, this flexibility to choose an optimal signal corresponding to a menu is reflected in the DM's attitude toward randomization. In a mixed menu $\alpha F + (1-\alpha)G$, the DM chooses acts $f \in F$ and $g \in G$ (i.e., a contingency plan $\alpha f + (1-\alpha)g$) not knowing whether the choice from F or G will actually determine her final outcome. Since the randomization realizes *after* the DM acquires information about the state ω, she is not able to tailor her information acquisition to the payoff relevant menu F or G. If the optimal signal differs for menus F and G, the DM would rather that the randomization over F and G is realized before she chooses what information to acquire. In particular, if the DM is indifferent between the menus F and G, she would prefer either one of them to the mixed menu $\alpha F + (1-\alpha)G$, where the randomization over F and G has not yet been resolved. Ergin and Saver (2010) call this behavior an *aversion to contingent planning*.[5]

Axiom 5 (Aversion to contingent planning). *For all menus F and G, if $F \sim G$, then $F \succsim \alpha F + (1-\alpha)G$ for all $\alpha \in (0,1)$.*

However, since information is redundant for singleton menus, the optimal information in a mixed menu $\alpha F + (1-\alpha)h$ depends only on α and F, and does not change if h is replaced by an the alternative act h'. As a result, the DM's preferences exhibit an *independence of degenerate decisions* (Ergin and Saver 2010).

Axiom 6 (Independence of degenerate decisions). *For all menus F and G, acts h and h', and $\alpha \in (0,1)$,*

$$\alpha F + (1-\alpha)h \succsim \alpha G + (1-\alpha)h \quad \Rightarrow \quad \alpha F + (1-\alpha)h' \succsim \alpha G + (1-\alpha)h'.$$

[4] A preference for flexibility distinguishes the information-acquisition problem in a rationally inattentive preference from models where the DM may be "inattentive" to some of the alternatives in a menu (see, e.g., Masatlioglu et al. 2012, Manzini and Mariotti 2014, or Ortoleva 2013).

[5] Axiom 5 could also be interpreted as expressing a desire for early resolution of uncertainty (Ergin and Sarver 2015). Moreover, Axiom 5 rules out a preference for hedging that, for instance, could arise if the DM today were uncertain about the information-acquisition technology available tomorrow. Indeed, replacing aversion to contingent planning with a desire for randomization (i.e., $F \sim G$ implies $\alpha F + (1-\alpha)G \succsim F$) would characterize a model in which the maximization in (1) is replaced with a minimization. Following Maccheroni et al. (2006), one could interpret such a model by saying that DM acts "as if" she is playing a game against a malevolent nature, where nature chooses a signal to minimize the benefits of information for each menu.

Finally, in the information-acquisition problem, the DM acquires information only about the objective state of fundamentals. As a result, her preferences over menus satisfy a state-by-state *dominance* axiom (Dillenberger et al. 2014). In particular, adding an act g to a menu F can make the DM strictly better off only if there is some information about the objective state that would lead the DM to choose g from $F \cup \{g\}$: if F already contains an act that is preferred to g in *all* states, adding g to her opportunity set cannot make her strictly better off.

Axiom 7 (Dominance). *For all menus F and acts g, if there exists $f \in F$ such that $f(\omega) \succsim g(\omega)$ for all $\omega \in \Omega$, then $F \sim F \cup \{g\}$.*

3.2 *Representation theorem*

The following theorem shows that Axioms 1–7 characterize all observable implications of rational inattentive preferences.

Theorem 1. *A binary relation $\succsim$ over menus is a rationally inattentive preference if and only if it satisfies Axioms 1–7.*

Theorem 1 shows that the information-acquisition problem in Section 2.2 induces intuitive behavioral traits that can be observed in our framework. It also establishes a formal connection between the literature on information acquisition (e.g., models of rational inattention) and the decision-theory literature on menu choice. Building on the seminal menu-choice paper by Kreps (1979), Dekel et al. (2001) model choice between menus of lotteries as that of a DM who exhibits a preference for flexibility because she expects to learn more about her taste before choosing a lottery from the menu. Ergin and Saver (2010) establish that relaxing the "set independence" axiom in Dekel et al. (2001) to aversion to contingent planning admits costly contemplation: the DM acts as if she is able to exert costly contemplation to reduce uncertainty about her future tastes before choosing a lottery from the menu. Dillenberger et al. (2014) introduce objective states into the menu-choice framework and model choice between menus of Anscombe–Aumann acts as that of a DM who expects to receive an exogenous signal about the state of the world before choosing an alternative from the menu. To focus on learning about the objective state, their characterization introduces a dominance axiom (closely related to our Axiom 7) in addition to axioms in Dekel et al. (2001).

Similar to the way that Ergin and Saver (2010) generalize the independence axiom in Dekel et al. (2001) to allow for costly contemplation (where learning about tastes is flexible), Theorem 1 relaxes the independence axiom in Dillenberger et al. (2014) to characterize costly information acquisition (where learning about objective states is flexible). While the axiomatic characterization shows that rationally inattentive preferences can be viewed as a class of costly contemplation preferences, our proof of Theorem 1 does not start from the representation theorem in Ergin and Saver (2010). In fact, the additional structure provided by the objective state space and Axiom 7 allows us to follow an alternative approach to the characterization, which uses methods developed in Maccheroni et al. (2006).

Maccheroni et al. (2006) pioneered the analysis of choice models with a variational structure in the context of ambiguity aversion. Our analysis is connected because the value of information, $b^u_F(\pi) - c(\pi)$, is concave and upper semicontinuous in π, and the optimization over signals therefore corresponds to a variational problem. Especially important for our approach to the proof of Theorem 1 is that, for rationally inattentive preferences, every menu F has a certainty equivalent (i.e., a constant act $x_f \in X$ such that $F \sim x_f$). For the general costly contemplation model in Ergin and Saver (2010), certainty equivalents do not follow from a natural assumption—such as Axiom 7—on preferences. The existence of certainty equivalents allows us to impose weaker assumptions on primitives (e.g., a weaker continuity axiom) and provide a proof that does not require the compactness properties Ergin and Saver's (2010) framework. An additional advantage is that our proof methods immediately extend to a more general framework with an infinite state space (see Appendix A.3).

3.3 *Special cases*

Special cases of rationally inattentive preferences can also be characterized in terms of the additional restrictions they impose on menu-choice data. For example, constrained-information acquisition can be characterized by an indifference toward contingency plans between menus and singletons.

Axiom 8 (Weak indifference to contingent planning). *For all menus F and acts h, $F \sim h$ implies $F \sim \alpha F + (1-\alpha)h$ for all $\alpha \in (0,1)$.*

In the general model characterized in Theorem 1, a DM is indifferent to randomizing over menus F and G if there is a common signal π that is optimal for both of these menus. A key feature of constrained information is that all signals in the constraint set $\Gamma(\bar{p})$ are costless, and so all of the signals in this set are optimal for a singleton menu. In particular, whatever signal is optimal for menu F is also optimal for menu h, and the DM is therefore indifferent to contingent planning with singletons. The following corollary shows that this additional axioms characterizes constrained-information preferences.

Corollary 1. *A binary relation $\succsim$ over menus is a constrained-information preference if and only if it satisfies Axioms 1–7 and 8.*

Alternatively, subjective learning can be characterized by indifference to contingent planning for arbitrary menus.

Axiom 5b (Indifference to contingent planning). *For all menus F and G, $F \sim G$ implies $F \sim \alpha F + (1-\alpha)G$ for all $\alpha \in (0,1)$.*

In the subjective-learning model, the DM always acquires the same information. In particular, she acquires exactly the same signal for any menus F and G, and this feature of the subjective-learning model is reflected in Axiom 5b. The following corollary shows that replacing Axiom 5 with Axiom 5b characterizes subjective-learning preferences.

COROLLARY 2. *A binary relation $\succsim$ over menus is a subjective-learning preference if and only if it satisfies Axioms 1–4, 5b, 6, and 7.*

Dillenberger et al. (2014) provide the first characterization of subjective-learning preferences and analyze this model in more detail.

4. IDENTIFYING PARAMETERS

In this section, our main objective is to show how the parameters $(u, \bar{p}, c)$ in the information-acquisition problem can be identified from menu-choice data. Identifying the utility and prior is straightforward. Our main result in this section (Theorem 2) shows that information costs can also be identified if we impose some additional properties to "normalize" the cost function.

4.1 *Identifying utility and prior*

The following remark first shows that it is straightforward to identify the utility u and prior $\bar{p}$ from the DM's preferences over singletons.

REMARK 1. Let $\succsim$ be a rationally inattentive preference represented by $(u, \bar{p}, c)$. Then (i) every menu F has a *certainty equivalent* $x_F \in X$ such that $F \sim x_F$ and (ii) for all acts f and g,

$$f \succsim g \quad \Leftrightarrow \quad \int_\Omega u\big(f(\omega)\big)\,\bar{p}(d\omega) \geq \int_\Omega u\big(g(\omega)\big)\,\bar{p}(d\omega).$$

Moreover, if $(u', \bar{p}', c')$ represents the same preference over menus, then $\bar{p}' = \bar{p}$ and there exist $\alpha > 0$ and $\beta \in \mathbb{R}$ such that $u' = \alpha u + \beta$.

The certainty equivalent of a menu in part (i) can be interpreted as the sure amount the DM would be willing to pay today to have F as her opportunity set tomorrow. In the classic Anscombe–Aumann framework, certainty equivalents for acts are used to identify a utility function u (representing the DM's risk preferences) and a prior $\bar{p}$ (representing the DM's initial beliefs). Part (ii) shows that, in our extension of the Anscombe–Aumann framework, preferences over singleton menus identify u and $\bar{p}$ in the same manner. In the following subsection, we therefore assume that u and $\bar{p}$ are given, and focus on how certainty equivalents for menus can be used to identify the DM's information costs.

4.2 *Identifying information costs*

In Definition 2, we impose only mild regularity conditions on information costs, which are necessary to ensure that the optimization over signals is well defined. In this general form, it is not possible to identify costs uniquely. For instance, assume that $(u, \bar{p}, c)$ represents a rationally inattentive preference. First, it is clearly possible to add a constant to the cost function c without changing the ordinal ranking over menus. Second,

if there are signals $\pi \trianglerighteq \rho$ such that $c(\pi) < c(\rho)$, then the DM would never choose signal ρ (because $b^u_F(\pi) \geq b^u_F(\rho)$ for all menus F), and so changing the value $c(\rho)$ to any number greater than $c(\pi)$ would lead to an alternative cost function that induces the same preferences. Finally, if for some signals π and ρ, and $\alpha \in (0,1)$, it is the case that $c(\alpha\pi + (1-\alpha)\rho) > \alpha c(\pi) + (1-\alpha)c(\rho)$, then the DM would never choose signal $\alpha\pi + (1-\alpha)\rho$ (because $b^u_F(\alpha\pi + (1-\alpha)\rho) = \alpha b^u_F(\pi) + (1-\alpha)b^u_F(\rho)$ for all menus F), so changing the value $c(\alpha\pi + (1-\alpha)\rho)$ to any number greater than $\alpha c(\pi) + (1-\alpha)c(\rho)$ would, again, lead to an alternative cost function that induces the same preferences.

To rule out these cases, we focus on the following class of *canonical* information costs.

Definition 3. An information cost function $c : \Pi(\bar{p}) \to [0,\infty]$ is canonical if the following properties are satisfied:

(i) No information is costless. We have $c(\pi_0) = 0$, where π_0 assigns probability 1 to the prior $\bar{p}$.

(ii) Blackwell monotonicity. We have that $\pi \trianglerighteq \rho$ implies $c(\pi) \geq c(\rho)$.

(iii) Convexity. For all signals π and ρ, and $\alpha \in (0,1)$,

$$c\big(\alpha\pi + (1-\alpha)\rho\big) \leq \alpha c(\pi) + (1-\alpha)c(\rho).$$

Properties (i)–(iii) are satisfied by most information cost functions used in applications. For example, it is well known that mutual information satisfies these properties (Cover and Thomas 2006, Chapter 2), and so the cost functions based on mutual information—which are frequently used in the rational inattention literature—are canonical.[6]

We can now state our identification result.

Theorem 2. *Let $\succsim$ be a rationally inattentive preference such that the restriction of $\succsim$ to singleton menus is represented by $(u, \bar{p})$. Then the cost function $c : \Pi(\bar{p}) \to [0,\infty]$, defined by*

$$c(\pi) = \sup_{F \in \mathbb{F}} \big[b^u_F(\pi) - u(x_F)\big], \tag{2}$$

is the unique canonical information cost function such that $(u, \bar{p}, c)$ represents $\succsim$.

As an immediate implication of Theorem 2 and Remark 1, properties (i)–(iii) allow an identification of all model parameters up to a standard positive affine transformation.

Corollary 3. *If $(u, \bar{p}, c)$ and $(u', \bar{p}', c')$ represent the same rationally inattentive preference $\succsim$, and c and c' are canonical, then there exists $\alpha > 0$ and $\beta \in \mathbb{R}$ such that $u' = \alpha u + \beta$, $\bar{p}' = \bar{p}$, and $c' = \alpha c$.*

[6]The linear cost function $c_\lambda(\pi)$ is an example of a cost function that is not only convex, but also linear in signals: $c_\lambda(\alpha\pi + (1-\alpha)\rho) = \alpha c_\lambda(\pi) + (1-\alpha)c_\lambda(\rho)$ for all signals π and ρ, and $\alpha \in (0,1)$.

Theorem 2 establishes that it is without loss of generality to focus on information costs satisfying properties (i)–(iii): *any* rationally inattentive preference $\succsim$ has a representation with canonical information costs. More importantly, there is always a unique information cost function satisfying properties (i)–(iii). As a result, it is not possible to further restrict the class of information costs, as any additional restriction would have additional behavioral implications. In that sense, Theorem 2 identifies a canonical class of information costs, which can be used to assess whether the implications of a particular model of information acquisition—such as the specific models used in the rational inattention literature—depend only on the general idea of costly information acquisition or on specific functional form assumptions.

Formula (2) for the information cost is also of practical relevance, because it shows how the canonical costs can be constructed from menu-choice data. Consider a rationally inattentive DM with preference relation $\succsim$ represented by $(u, \bar{p}, c)$. By Remark 1, the restriction of $\succsim$ to singleton menus identifies $(u, \bar{p})$, and so preferences over singletons can be used to construct $(u, \bar{p})$ in the standard way. Now suppose the objective is to measure the cost the DM incurs to extract a particular signal $\pi \in \Pi(\bar{p})$. Since a rationally inattentive DM chooses signals optimally, we know that for any menu F, $V(F) \geq b_F^u(\pi) - c(\pi)$ (where V is the preference functional in Definition 2). Using the certainty equivalent in Remark 1, this inequality gives a lower bound of $b_F^u(\pi) - u(x_F)$ on the information cost $c(\pi)$. By varying the menu F, one tightens the lower bound and, therefore, achieves an increasingly precise estimate of the DM's information cost for signal π. Theorem 2 shows exactly what these lower bounds are approximating: it is the unique canonical information cost.[7]

Identification results related to Theorem 2 are given for variational preferences in other settings by Maccheroni et al. (2006) (in the context of ambiguity aversion) and Ergin and Saver (2010) (for costly contemplation). However, Theorem 2 is conceptually and technically different from these existing identification results. Maccheroni et al. (2006) study variational preferences over Anscombe–Aumann acts and show that unboundedness of the utility index is sufficient to identify a unique cost function—interpreted as an ambiguity index in their model—that is grounded and convex. Ergin and Saver (2010) restrict the set of contemplation strategies to a "minimal" set. This identification approach is natural when the state space is subjective (see, e.g., Dekel et al. 2001), but is less appealing when there are objective states and the set of signals $\Pi(\bar{p})$ is objectively given. From a practical point of view, it is also more straightforward to verify properties of a cost function than to check that a set of signals is minimal.

We therefore take a different approach to the identification problem using the additional structure in our framework and adapting an argument in Sarver (2014). In particular, a distinctive feature of our analysis is the interaction between the variational nature of preferences and the Blackwell order, which plays a crucial role in Theorem 2 and has no counterpart in Maccheroni et al. (2006) or Ergin and Saver (2010).

[7] Caplin and Dean (2014), Gabaix et al. (2006), and Cheremukhin et al. (2015) propose experimental designs to study rational inattention that could be adapted to measure hidden information costs using the insights from Theorem 2.

To illustrate the role of the Blackwell order, we provide a brief outline of the proof of Theorem 2. Fix some parameters $(u, \bar{p}, c)$ representing a rationally inattentive preference $\succsim$. The first step of the proof shows that it is without loss of generality to assume that c is canonical. The second step is to show that if c is canonical, it is uniquely identified by formula (2). For this part of the proof, we first identify each menu F with a support function φ_F, which (in our context) is defined on the set of posteriors:

$$\varphi_F(p) = \max_{f \in F} \int_\Omega u\big(f(\omega)\big)p(d\omega) \quad \forall p \in \Delta(\Omega).$$

Support functions φ_F are continuous and convex, and the DM's preference over menus can be associated with a value function V over support functions by

$$V(\varphi_F) = \max_{\pi \in \Pi(\bar{p})} \int_{\Delta(\Omega)} \varphi_F(p)\pi(dp) - c(\pi).$$

This value function can be extended to the set of all continuous functions $\varphi : \Delta(\Omega) \to \mathbb{R}$ in the obvious way, and since c is convex and lower semicontinuous, an application of the Fenchel–Moreau theorem gives that

$$c(\pi) = \sup_{\varphi} \int_{\Delta(\Omega)} \varphi(p)\pi(dp) - V(\varphi) \quad \forall \pi \in \Pi(\bar{p}),$$

where the supremum is taken over all continuous functions $\varphi : \Delta(\Omega) \to \mathbb{R}$. The core of the proof (a separation argument) uses Blackwell monotonicity to show that the same supremum is achieved if we consider only support functions φ_F corresponding to the menus F in our framework, thereby establishing (2) and connecting the canonical information cost function directly with observable menu-choice data.

4.3 *Comparing information costs*

Finally, the uniqueness of canonical information costs allows us to ask how a change in costs translates into observable changes in behavior. To formalize such comparative statics, consider two DMs, DM1 and DM2, with rationally inattentive preferences $\succsim_1$ and $\succsim_2$ represented by $(u_1, \bar{p}_1, c_1)$ and $(u_2, \bar{p}_2, c_2)$, respectively, where c_1 and c_2 are canonical. To isolate the affect of information costs, we assume a common prior and utility function, and say that DM1 has *lower information costs* than DM2 if $(u_1, \bar{p}_1) = (u_2, \bar{p}_2)$ and $c_1 \leq c_2$. The following corollary shows how lower information costs are revealed by observable menu-choice behavior.

Corollary 4. *The following conditions are equivalent:*

(i) DM1 has lower information costs than DM2.

(ii) For all menus F and acts h,

$$h \succ_1 F \quad \Rightarrow \quad h \succ_2 F.$$

For a binary relation $\succsim$, $h \succ F$ implies that the flexibility of choosing an act $f \in F$ is not desirable enough for the DM to prefer menu F to the singleton menu h. Part (ii) therefore indicates that DM1 has a *stronger desire for flexibility* than DM2 (Dekel et al. 2001, Ergin and Saver 2010, and Dillenberger et al. 2014). As a result, Corollary 4 provides a behavioral interpretation for measures of information commonly used in applications. For instance, in the context of Example 2.3, a stronger desire for flexibility reveals when DM1 has a lower unit cost of information (λ) or a higher capacity constraint (κ).

5. Information acquisition

In this section, we further demonstrate the connection between menu choice and models of costly information acquisition by showing how a DM's preferences reveal the amount of information she acquires, an important question in applications.

To formalize this question, we observe that, in addition to a preference relation over menus (Definition 2), the information-acquisition problem with parameters $(u, \bar{p}, c)$ induces a choice correspondence $\mathcal{C} : \mathbb{F} \rightrightarrows \Pi(\bar{p})$ over signals:

$$\mathcal{C}(F) = \arg\max_{\pi \in \Pi(\bar{p})} \big[b^u_F(\pi) - c(\pi) \big].$$

The choice correspondence $\mathcal{C}$ is of primary interest in many applications, but is generally not directly observable. For example, in models of rational inattention, the signal an agent acquires is interpreted as a reflecting her choice about what information to pay attention to, which depends on hidden information costs. Alternative choice data must therefore be used to evaluate when one DM acquires more information than another.[8]

In the following discussion, we fix $(u, \bar{p})$ and consider two canonical information cost functions c_1 and c_2. Denote by $\succsim_i$ the preference relation over menus and denote by $\mathcal{C}_i$ the choice correspondence over signals, induced by $(u, \bar{p}, c_i)$ for $i = 1, 2$. Intuitively, given two DMs (with cost functions c_1 and c_2), we want to say that DM1 acquires more information than DM2 if she acquires Blackwell more informative signals for every menu. Since $\mathcal{C}_i(F)$ may not be a singleton, formalizing this idea requires an extension of the Blackwell order to sets of signals.

Definition 4. The set of signals Π_1 is *more informative* than Π_2, denoted $\Pi_1 \trianglerighteq \Pi_2$, if

$$\sup_{\pi \in \Pi_1} \int_{\Delta(\Omega)} \varphi(p)\pi(dp) \geq \sup_{\rho \in \Pi_2} \int_{\Delta(\Omega)} \varphi(p)\rho(dp)$$

for every convex continuous functions $\varphi : \Delta(\Omega) \to \mathbb{R}$.

Definition 4 is a natural generalization of the Blackwell order to sets of signals: if Π_1 is more informative than Π_2, a DM with any payoff function φ over posteriors would

[8] Caplin and Dean (2015) analyze how this can be done by looking at ex post data, i.e., the stochastic state-dependence choice of acts from menus after the decision-maker acquires signals. We show below that this can also be done with menu choices, which provides an alternative data set that may be useful when state-dependent stochastic choice data are not available.

prefer to have access to the set of signals Π_1 than the set of signals in Π_2 (in particular, the restriction to singleton sets coincides with the usual Blackwell order on signals, explaining our abuse of notation).

We use this extension of the Blackwell order to sets of signals to formally define when DM1 *acquires more information than* DM2.

DEFINITION 5. DM1 acquires more information than DM2 if $\mathcal{C}_1(F) \trianglerighteq \mathcal{C}_2(F)$ for all menus F.

Our objective is to translate the above condition into observable choice behavior. To do so, we start from the intuition that a DM who anticipates being more informed will have a stronger preference for menus that offer a higher premium for information (i.e., menus for which information is more valuable). Of course, in general, DMs can disagree about the information premium offered by different menus, so we must first identify pairs of menus where all DMs should agree about which one has the higher information premium. Since information is not valuable for singleton menus, we start by observing that a singleton menu h has a lower information premium than any other menu G. Now suppose a coin is tossed, so that the final choice of an act will come from a common menu F with probability α and either menu G or h otherwise. In the menu $\alpha F + (1-\alpha)h$, information has value only if the final choice comes from F, while in the menu $\alpha F + (1-\alpha)G$, the same information may allow the DM to choose a better alternative from both menu F and menu G. Thus, we argue that any DM should regard menu $\alpha F + (1-\alpha)G$ as offering a higher *information premium* than menu $\alpha F + (1-\alpha)h$. Accordingly, the following Theorem 3 shows that DM1 acquires more information than DM2 if and only if whenever DM2 prefers a menu with a higher information premium, DM1 does likewise.

THEOREM 3. *The following conditions are equivalent:*

(i) DM1 acquire more information than DM2.

(ii) For all menus F and G, acts h, and $\alpha \in (0,1)$,

$$\alpha F + (1-\alpha)h \succ_1 \alpha F + (1-\alpha)G \quad \Rightarrow \quad \alpha F + (1-\alpha)h \succ_2 \alpha F + (1-\alpha)G.$$

Theorem 3 translates comparative statics on unobservable signal choices (how much information DMs acquire) into a comparative statics criterion on observable choice data (preferences for an information premium). In particular, part (ii) can be interpreted in terms of the primitive behavioral traits in Section 3. While it is sufficient to consider the DMs' desires for flexibility to compare their information costs (Corollary 4), comparing how much information the DMs acquire depends on both their desire for flexibility and their aversion to contingency planning. Consider a DM comparing menus $\alpha F + (1-\alpha)G$ and $\alpha F + (1-\alpha)h$. On one hand, menu G offers more flexibility to adjust to new information than the singleton menu h. On the other hand, the randomization in mixed menu $\alpha F + (1-\alpha)G$ does not allow the DM to target her choice of a signal specifically to the menu that determines her final payoffs. As such, the DM

faces a trade-off between her preference for flexibility and her aversion to contingency planning. Theorem 3 therefore reflects the intuition that if DM1 anticipates that she will acquire more information than DM2, she is better able to exploit the flexibility offered by menu $\alpha F + (1-\alpha)G$ and is also less affected by the need to make contingency plans.

6. Conclusion

In this paper, we show how menu-choice data can be used to study models of costly information acquisition with hidden information costs. Such models have recently gained prominence with the rational inattention literature (Sims 1998, 2003), where costs of information are interpreted as representing limitations on attention and are therefore not directly observable. We complement applied research on rational inattention by providing a theory for how such models of individual behavior can be tested, and how hidden information costs can be identified and elicited with observable choice data.

Our framework takes the perspective of an agent when she chooses a menu—an opportunity set—with the prospect of acquiring an informative signal about the state of the world before selecting an alternative. Preferences over menus reveal the trade-off the DM faces when she balances the benefits of information against the cost of signals that convey information about the state. We show that three properties—(i) acquiring no information is costless, (ii) Blackwell monotonicity, and (iii) convexity—are sufficient to identify information costs from menu-choice data. Moreover, an explicit formula relates the unique canonical information cost to the DM's willingness to pay for different menus, suggesting a specific procedure by which hidden information costs can be elicited from data in dynamic choice environments or experimental settings. We also provide a theory of comparative statics, which shows how menu-choice data can be used to compare information costs, and reveal how much information rationally inattentive agents acquire.

Our analysis also provides a natural starting point for research on the behavioral foundations of dynamic models of rational inattention. Extending on our analysis, future research in these direction has the potential to support much needed empirical analysis of the implications of costly information acquisition in a variety of economic environments.

Appendix

In this section, we prove the results in Sections 3–5.

A.1 *Preliminaries*

We first introduce some additional notation and preliminary results required for the proofs.

Niveloids Denote by $C(\Delta(\Omega))$ the linear space of real-valued continuous functions defined on $\Delta(\Omega)$, and denote by $\mathrm{ca}(\Delta(\Omega))$ the linear space of signed measures of bounded

variation on $\Delta(\Omega)$ (Aliprantis and Border 2006, p. 399). For each $\pi \in \mathrm{ca}(\Delta(\Omega))$ and for each $\varphi \in C(\Delta(\Omega))$, let

$$\langle \varphi, \pi \rangle = \int_{\Delta(\Omega)} \varphi(p)\pi(dp).$$

The linear space $C(\Delta(\Omega))$ is endowed with the supnorm and $\mathrm{ca}(\Delta(\Omega))$ is endowed with the weak* topology. Therefore $\mathrm{ca}(\Delta(\Omega))$ can be identified with the continuous dual space of $C(\Delta(\Omega))$ (Aliprantis and Border 2006, Corollary 14.15), and $C(\Delta(\Omega))$ can be identified with the continuous dual space of $\mathrm{ca}(\Delta(\Omega))$ (Aliprantis and Border 2006, Theorem 5.93).

Let Ψ be a subset of $C(\Delta(\Omega))$ and consider a function $V : \Psi \to \mathbb{R}$. We say that V is *normalized* if $V(\alpha) = \alpha$ for each constant function $\alpha \in \Psi$, is *monotone* if $V(\varphi) \geq V(\psi)$ for all $\varphi, \psi \in \Psi$ such that $\varphi \geq \psi$, is *translation invariant* if $V(\varphi + \alpha) = V(\varphi) + \alpha$ for each $\varphi \in \Psi$ and $\alpha \in \mathbb{R}$ such that $\varphi + \alpha \in \Psi$, and is a *niveloid* if $V(\varphi) - V(\psi) \leq \sup\{\varphi(p) - \psi(p) : p \in \Delta(\Omega)\}$ for each $\varphi, \psi \in \Psi$.

Niveloids are studied in detail in Cerreia-Vioglio et al. (2014), who prove also the following results. If V is a niveloid, then it is monotone and translation invariant, while the converse is true whenever $\Psi = \Psi + \mathbb{R}$. Moreover, if V is a niveloid, then V is (Lipschitz) continuous. If Ψ is a convex set and V is a convex niveloid, then there is a convex niveloid that extends V to $C(\Delta(\Omega))$.

Notation and auxiliary results Let Φ be the set of convex functions belonging to $C(\Delta(\Omega))$: Φ is a closed convex cone such that $0 \in \Phi$. Denote by Φ^* the dual cone of Φ, that is,

$$\Phi^* = \big\{\pi \in \mathrm{ca}\big(\Delta(\Omega)\big) : \langle \varphi, \pi \rangle \geq 0 \text{ for all } \varphi \in \Phi\big\}.$$

The set Φ^* is also a closed convex cone such that $0 \in \Phi^*$. Moreover $\Phi = \Phi^{**}$ (see Aliprantis and Border 2006, Theorem 5.103), that is,

$$\Phi = \big\{\varphi \in C\big(\Delta(\Omega)\big) : \langle \varphi, \pi \rangle \geq 0 \text{ for all } \pi \in \Phi^*\big\}.$$

Let $u : X \to \mathbb{R}$ be an affine function. Denote by $\Phi_{\mathbb{F}}$ ($\Phi_{\mathcal{F}}$, Φ_X) the set of functions $\varphi_F : \Delta(\Omega) \to \mathbb{R}$ ($\varphi_f : \Delta(\Omega) \to \mathbb{R}$, $\varphi_x : \Delta(\Omega) \to \mathbb{R}$) such that for some menu F (act f, outcome x),

$$\varphi_F(p) = \max_{f \in F} \int_\Omega u\big(f(\omega)\big)p(d\omega) \quad \left(\varphi_f(p) = \int_\Omega u\big(f(\omega)\big)p(d\omega), \varphi_x(p) = u(x)\right)$$

for all $p \in \Delta(\Omega)$.

Observe that $u(X) = \Phi_X \subset \Phi_{\mathcal{F}} \subset \Phi_{\mathbb{F}} \subset \Phi$. Moreover, $\alpha\varphi_F + (1-\alpha)\varphi_G = \varphi_{\alpha F + (1-\alpha)G}$ for each pair of menus F and G, and $\alpha \in [0,1]$. Hence, in particular, $\Phi_{\mathbb{F}}$ is convex.

Recall that, for a rationally inattentive preference, $u(X)$ is unbounded. The following additional properties of $\Phi_{\mathcal{F}}$ and $\Phi_{\mathbb{F}}$ hold when $u(X)$ is unbounded above (analogous properties hold when $u(X)$ is unbounded below):

(i) We have $\Phi_{\mathbb{F}} + [0, \infty) = \Phi_{\mathbb{F}}$.

(ii) The inequality $\varphi_F \geq 0$ implies $\alpha\varphi_F \in \Phi_{\mathbb{F}}$ for every $\alpha \in [1, \infty)$.

(iii) The relation $0 \in u(X)$ implies $\alpha\varphi_F \in \Phi_{\mathbb{F}}$ for every $\alpha \in [0,1)$.

(iv) If $u(X)$ is open, for each menu F, there is $\alpha > 1$ such that $\alpha\varphi_F \in \Phi_{\mathbb{F}}$.

(v) The term $\Phi_{\mathbb{F}} + \mathbb{R}$ is dense in Φ.

Properties (i)–(iv) are easily verified. To show that property (v) holds, let $\varphi \in \Phi$. By adding a constant to φ, we can assume without loss of generality that the image of φ lies in the interior of $u(X)$. By Theorem 1.7.1 in Schneider (1993), there exists a compact, convex set $K \subset u(X)^{\Omega}$ such that

$$\varphi(p) = \max_{k \in K} \int_{\Omega} k(\omega)\, p(d\omega).$$

By Theorem 1.8.16 in Schneider (1993), for any $\epsilon > 0$, there exists a finite set $H \subset \mathbb{R}^{\Omega}$ such that $H \subset K \subset \operatorname{co}(H) + \epsilon B$, where B is the unit ball in $\mathbb{R}^{\Omega}$. For ϵ sufficiently small, there exists a menu $F \in \mathbb{F}$ such that $H = u(F)$, and so $\varphi_F \leq \varphi \leq \varphi_F + \epsilon$.

A.2 *Proofs*

For simplicity, we assume that utility functions are unbounded above (the case where they are unbounded below is analogous and is omitted).

Proof of Theorem 1 It is straightforward to show that a rationally inattentive preference $\succsim$ satisfies Axioms 1–7, so we omit this direction of the proof. To prove the converse, we start by establishing an implication of Axioms 4 and 7 that we will use throughout the proof.

Claim 1. *For menus F and G, suppose that for each $g \in G$, there is $f \in F$ such that $f(\omega) \succsim g(\omega)$ for all ω. Then $F \succsim G$.*

Proof. Let $F = \{f_1, \ldots, f_n\}$ and $G = \{g_1, \ldots, g_m\}$. By Axiom 4,

$$F \cup G \succsim \cdots \succsim \{f_1, f_2\} \cup G \succsim \{f_1\} \cup G \succsim G.$$

By Axiom 7,

$$F \sim F \cup \{g_1\} \sim F \cup \{g_1, g_2\} \sim \cdots \sim G \cup F.$$

Hence, we conclude that $F \succsim G$, as wanted. □

Claim 2. *Every menu F has a certainty equivalent $x_F \in X$ such that $x_F \sim F$.*

Proof. Since F and Ω are finite, we can let x be a best outcome and let y be a worst outcome that may occur in any act in F. By Claim 1 we have $x \succsim F \succsim y$. Now consider the two sets

$$A = \{\alpha \in [0,1] : \alpha x + (1-\alpha)y \succsim F\} \quad \text{and} \quad B = \{\alpha \in [0,1] : F \succsim \alpha x + (1-\alpha)y\}.$$

Then $A \cup B = [0,1]$ and, by Axiom 2, A and B are closed. Since $[0,1]$ is connected, there exists $\alpha \in A \cap B$ such that $\alpha x + (1-\alpha)y \sim F$. So let x_F be equal to $\alpha x + (1-\alpha)y$. □

Claim 3. *There exist an affine utility function $u : X \to \mathbb{R}$ with unbounded range and a prior probability measure $\bar{p}$ over Ω such that the preference $\succsim$ over $\mathcal{F}$ is represented by the function $U : \mathcal{F} \to \mathbb{R}$ defined by*

$$U(f) = \int_\Omega u\big(f(\omega)\big)\bar{p}(d\omega) \quad \forall f \in \mathcal{F}.$$

Proof. Take $f, g \in \mathcal{F}$ and fix $\alpha \in [0,1]$. Suppose that $f \sim g$. By Axiom 5, $f = \alpha f + (1-\alpha)f \succsim \alpha g + (1-\alpha)f$. By Axiom 6, $\alpha f + (1-\alpha)g \succsim \alpha g + (1-\alpha)g = g$. By Axiom 5, $g = \alpha g + (1-\alpha)g \succsim \alpha f + (1-\alpha)g$. So we conclude that $f \sim g \sim \alpha f + (1-\alpha)g$. The remainder of the proof then follows from Maccheroni et al. (2006, Corollary 20 and Lemma 29). □

It is without loss of generality to let $0 \in u(X)$ and assume that $u(x) \geq 0$ for each $x \in X$ whenever $u(X)$ is lower bounded and closed.

Now, with some abuse of notation, define the functional $V : \Phi_{\mathbb{F}} \to \mathbb{R}$ such that $V(\varphi_F) = U(x_F)$, where x_F is a certainty equivalent of F.[9] If x_F and y_F are two certainty equivalents of F, then $x_F \sim y_F$ and so $U(x_F) = U(y_F)$. To conclude that V is well defined, we need to show that $\varphi_F = \varphi_G$ implies $F \sim G$ for each pair of menus F and G. The next two claims will accomplish this goal.

Claim 4. *Consider a pair of menus F and G. If $\varphi_F \geq \varphi_G$, then for each $g \in G$, there exists $f \in \operatorname{co} F$ (where $\operatorname{co} F$ is the convex hull of F) such that $f(\omega) \succsim g(\omega)$ for each $\omega \in \Omega$.*

Proof. We prove the contrapositive. Assume that there is $g \in G$ such that for all $f \in \operatorname{co} F$, we have $g(\omega) \succ f(\omega)$ for some $\omega \in \Omega$. Consider $u(\operatorname{co} F)$. By affinity of u, $\operatorname{co}(u(F)) = u \circ (\operatorname{co} F)$, so that $u(\operatorname{co} F)$ is convex, closed, and bounded. Let $E = \{e \in \mathbb{R}^\Omega : e \geq u \circ g\}$. Then E is closed convex cone. Clearly, $u(\operatorname{co} F)$ and E are disjoint. By a separating hyperplane theorem (Rockafellar 1970, Corollary 11.4.2), there exists some $p \in \mathbb{R}^\Omega$ such that

$$\int_\Omega u\big(f(\omega)\big)p(d\omega) < \int_\Omega e(\omega)p(d\omega) \quad \forall e \in E \text{ and } \forall f \in F.$$

Since $u \circ g$ belongs to E, we have

$$\max_{f \in F} \int_\Omega u\big(f(\omega)\big)p(d\omega) < \int_\Omega u\big(g(\omega)\big)p(d\omega).$$

Hence, since E is a cone, it is possible to choose $p \in \Delta(\Omega)$ so that $\varphi_F(p) < \varphi_G(p)$. □

Claim 5. *Consider a pair of menus F and G. If $G \subset \operatorname{co} F$, then $F \succsim G$.*

[9] For convenience we use V to denote both the representation over menus and the induced representation over support functions.

PROOF. Let $G = \{g_1, \dots, g_n\} \subset \operatorname{co} F$. For all $i = 1, \dots, n$, we can write each $g_i = \sum_{j=1}^{m_i} \alpha_j^i f_j^i$ for $\alpha_1^i, \dots, \alpha_{m_i}^i \geq 0$ summing up to 1, and $f_1^i, \dots, f_{m_i}^i \in F$. Hence

$$G \subset \sum_{j=1}^{m_1} \cdots \sum_{j'=1}^{m_n} \alpha_j^1 \cdots \alpha_{j'}^n F = \sum_{k=1}^{l} \beta_k F.$$

By Axiom 4 we have that $\sum_{k=1}^{l} \beta_k F \succsim G$, so it is enough to check that $F \sim \sum_{k=1}^{l} \beta_k F$. We show this by induction on l. If $l = 1$, then $\sum_{k=1}^{l} \beta_k F = F \sim F$. Suppose now the claim is true for $l - 1$. Observe that

$$\sum_{k=1}^{l} \beta_k F = \beta_l F + (1 - \beta_l) \left(\sum_{k=1}^{l-1} \frac{\beta_k}{1 - \beta_l} F \right).$$

Moreover, by inductive assumption, $F \sim \sum_{k=1}^{l-1} (\beta_k / (1 - \beta_l)) F$. Therefore, by Axiom 5, $F \succsim \sum_{k=1}^{l} \beta_k F$. Since $F \subset \sum_{k=1}^{l} \beta_k F$, by Axiom 4 we obtain $\sum_{k=1}^{l} \beta_k F \succsim F$. Therefore, $F \sim \sum_{k=1}^{l} \beta_k F$, as wanted. □

By Claim 4, if $\varphi_F \geq \varphi_G$, then there exists a subset $H \subset \operatorname{co} F$ such that for each $g \in G$ there exists $h \in H$ such that $h(\omega) \succsim g(\omega)$ for all $\omega \in \Omega$. By Claim 5, F is preferred to H, which, by Claim 1, is preferred to G. This shows that V is well defined (and monotone). Moreover, notice that V represents $\succsim$ in the sense that $F \succsim G$ if and only if $V(\varphi_F) \geq V(\varphi_G)$.

CLAIM 6. *The functional V is a monotone, normalized, convex niveloid.*

PROOF. The monotonicity of V comes immediately from Claims 4 and 5. Moreover, observe that the set of constant functions in $\Phi_{\mathbb{F}}$ is Φ_X, and for every outcome x, we have $V(\varphi_x) = u(x) = \varphi_x$, so that V is normalized.

We next show that V is translation invariant. Using Axiom 6, the obvious adaptation of the argument in Maccheroni et al. (2006, Proof of Lemma 28) provides that whenever k belongs to $u(X)$, we have for any $\varphi_F \in \Phi_{\mathbb{F}}$,

$$V\big(\beta\varphi_F + (1 - \beta)k\big) = V(\beta\varphi_F) + (1 - \beta)k \quad \forall \beta \in (0, 1).$$

Pick $\gamma > 1$, so that $\gamma\varphi_F \in \Phi_{\mathbb{F}}$. Then

$$V\left(\frac{1}{\gamma}(\gamma\varphi_F) + \frac{\gamma - 1}{\gamma}\left(\frac{\gamma}{\gamma - 1}k\right)\right) = V\left(\frac{1}{\gamma}(\gamma\varphi_F)\right) + \frac{\gamma - 1}{\gamma}\left(\frac{\gamma}{\gamma - 1}k\right) \quad \forall \alpha > 0.$$

This implies that $V(\varphi_F + k) = V(\varphi_F) + k$ whenever $k > 0$. Now fix $\varphi_F \in \Phi_{\mathbb{F}}$. For any $k < 0$ such that $\varphi_F + k \in \Phi_{\mathbb{F}}$,

$$V(\varphi_F) = V(\varphi_F + k - k) = V(\varphi_F + k) - k \quad \Rightarrow \quad V(\varphi_F) + k = V(\varphi_F + k).$$

Hence, V is translation invariant on $\Phi_{\mathbb{F}}$.

To show that V is convex, suppose $V(\varphi_F) = V(\varphi_G)$. Then $F \sim G$ and, by Axiom 5, $F \succsim \alpha F + (1-\alpha)G$. Hence,

$$\alpha V(\varphi_F) + (1-\alpha)V(\varphi_G) = V(\varphi_F) \geq V(\varphi_{\alpha F+(1-\alpha)G}) = V\big(\alpha\varphi_F + (1-\alpha)\varphi_G\big).$$

Now suppose $V(\varphi_G) > V(\varphi_F)$ and define $\beta = V(\varphi_G) - V(\varphi_F) > 0$. Since $\varphi_F + \beta \in \Phi_{\mathbb{F}}$,

$$V(\varphi_F + \beta) = V(\varphi_F) + \beta = V(\varphi_F) + V(\varphi_G) - V(\varphi_F) = V(\varphi_G),$$

where the first equality holds by translation invariance. Therefore,

$$\begin{aligned} V(\varphi_G) &\geq V\big(\alpha(\varphi_F + \beta) + (1-\alpha)\varphi_G\big) = V\big(\alpha\varphi_F + (1-\alpha)\varphi_G\big) + \alpha\beta \\ &= V\big(\alpha\varphi_F + (1-\alpha)\varphi_G\big) + \alpha\big(V(\varphi_G) - V(\varphi_F)\big), \end{aligned}$$

so that $V(\alpha\varphi_F + (1-\alpha)\varphi_G) \leq \alpha V(\varphi_F) + (1-\alpha)V(\varphi_G)$, and V is convex.

Since V is translation invariant on $\Phi_{\mathbb{F}}$, we can extend V uniquely to $\Phi_{\mathbb{F}} + \mathbb{R}$ by defining $V(\varphi) = V(\varphi + k) - k$ for any $\varphi \in \Phi_{\mathbb{F}} + \mathbb{R}$ and $k \in \mathbb{R}$ such that $\varphi + k \in \Phi_{\mathbb{F}}$. This extension preserves not only translation invariance, but also monotonicity and convexity. Hence the extension of V is a convex niveloid on $\Phi_{\mathbb{F}} + \mathbb{R}$ and a fortiori on $\Phi_{\mathbb{F}}$.

To complete the proof we apply the well known Fenchel–Moreau theorem (adapted to our framework). □

Claim 7. *There exists a proper, lower-semicontinuous cost function $c : \Pi(\bar{p}) \to [0, \infty]$ such that*

$$V(\varphi_F) = \max_{\pi \in \Pi(\bar{p})} \langle \varphi_F, \pi \rangle - c(\pi) \quad \forall F \in \mathbb{F}.$$

Proof. Since $\Phi_{\mathbb{F}}$ is convex and V is a convex niveloid, there is a real-valued functional W defined on $C(\Delta(\Omega))$ that is a convex niveloid extending V (see Section A.1). Since W is a niveloid, it is continuous. Since W is continuous, convex, and real-valued, by Rockafellar (1974, Theorem 11) the subdifferential of W is nonempty at each $\varphi \in C(\Delta(\Omega))$, that is, for each φ there is $\pi \in \mathrm{ca}(\Delta(\Omega))$ such that

$$\langle \varphi, \pi \rangle - W(\varphi) \geq \langle \psi, \pi \rangle - W(\psi) \quad \forall \psi \in C\big(\Delta(\Omega)\big). \tag{3}$$

Moreover, since W is a niveloid, it is monotone and translation invariant, so by Ruszczyński and Shapiro (2006, Theorem 2.2) we can choose π to be in $\Delta(\Delta(\Omega))$. Define $V^* : \Delta(\Delta(\Omega)) \to (-\infty, \infty]$ such that

$$V^*(\pi) = \sup_{F \in \mathbb{F}} \langle \varphi_F, \pi \rangle - V(\varphi_F) \quad \forall \pi \in \Delta\big(\Delta(\Omega)\big).$$

Thus, for all φ_F and π, $V^*(\pi) \geq \langle \varphi_F, \pi \rangle - V(\varphi_F)$ and hence $V(\varphi_F) \geq \langle \varphi_F, \pi \rangle - V^*(\pi)$. Moreover, by (3), for any φ_F, there exists a $\pi \in \Delta(\Delta(\Omega))$ such that $\langle \varphi_F, \pi \rangle - V(\varphi_F) = V^*(\pi)$. As a result,

$$V(\varphi_F) = \max_{\pi \in \Delta(\Delta(\Omega))} \langle \varphi_F, \pi \rangle - V^*(\pi) \quad \forall F \in \mathbb{F}.$$

We want c to be the restriction of V^* to $\Pi(\bar{p})$. To do so, we need to show that (i) $V^* \geq 0$, (ii) V^* is proper and lower semicontinuous, and (iii) $V^*(\pi) < \infty$ implies $\pi \in \Pi(\bar{p})$. First, since V is normalized and $0 \in \Phi_{\mathbb{F}}$, $\langle 0, \pi \rangle - V(0) = 0$, it follows that $V^* \geq 0$. Next, observe that

$$\langle \varphi_F, \bar{p} \rangle = \max_{f \in F} \int_\Omega u\big(f(\omega)\big)\bar{p}(d\omega) \quad \forall F \in \mathbb{F}.$$

By Axiom 4, F is preferred to any $f \in F$. Therefore, $V^*(\pi_0) \leq 0$ and hence $V^*(\pi_0) = 0$. This implies that V^* is proper. Lower semicontinuity comes from the fact that V^* is the pointwise supremum of a family of continuous functions. Finally, suppose that $V^*(\pi) < \infty$. For each $n \in \mathbb{N}$, choose consequences x and y such that $u(x) = n$ and $u(y) = 0$. Fix $\omega \in \Omega$ and consider an act f taking value x on ω and y otherwise. Then

$$\langle \varphi_f, \pi \rangle - V^*(\pi) = n \int_{\Delta(\Omega)} p(\omega)\pi(dp) - V^*(\pi) \leq V(\varphi_f) = n\bar{p}(\omega).$$

Since the above inequality holds for each n, as long as $V^*(\pi) < \infty$, it follows that

$$\int_{\Delta(\Omega)} p(\omega)\pi(dp) \leq \bar{p}(\omega) \quad \forall \omega \in \Omega,$$

and so, since $\int_{\Delta(\Omega)} p\pi(dp) \in \Delta(\Omega)$, it follows that $\int_{\Delta(\Omega)} p(\omega)\pi(dp) = \bar{p}(\omega)$ for all $\omega \in \Omega$. Hence,

$$V(\varphi_F) = \max_{\pi \in \Pi(\bar{p})} \langle \varphi_F, \pi \rangle - V^*(\pi) \quad \forall F \in \mathbb{F},$$

and we can let c be the restriction of V^* to $\Pi(\bar{p})$, as desired. □

The following lemma, which isused in the proof of Corollary 2, shows when the DM will choose different signals for different menus.

LEMMA 1. *Let $\succsim$ be a rationally inattentive preference represented by $V : \mathbb{F} \to \mathbb{R}$. Then for all finite collection of menus $F_1, \ldots, F_n$, and $\alpha_1, \ldots, \alpha_n > 0$ summing to 1, the following statements are equivalent:*

(i) We have $\alpha_1 V(F_1) + \cdots + \alpha_n V(F_n) = V(\alpha_1 F_1 + \cdots + \alpha_n F_n)$.

(ii) We have $\mathcal{C}(\alpha_1 F_1 + \cdots + \alpha_n F_n) \subset \mathcal{C}(F_i)$ for all $i = 1, \ldots, n$.

(iii) We have $\mathcal{C}(F_1) \cap \cdots \cap \mathcal{C}(F_n) \neq \varnothing$.

PROOF. We first show that (i) implies (ii) by induction on n. If $n = 1$, (i) trivially implies (ii). Now suppose that this is true also for $n - 1$. Set

$$G = \frac{\alpha_2}{1 - \alpha_1} F_2 + \cdots + \frac{\alpha_n}{1 - \alpha_1} F_n.$$

Notice that since V is convex, $\alpha_2 V(F_2) + \cdots + \alpha_n V(F_n) \geq (1 - \alpha_1)V(G)$ and

$$\alpha_1 V(F_1) + \cdots + \alpha_n V(F_n) = V\big(\alpha_1 F_1 + (1 - \alpha_1)G\big) \leq \alpha_1 V(F_1) + (1 - \alpha_1)V(G).$$

Therefore, $\alpha_2 V(F_2) + \cdots + \alpha_n V(F_n) = (1-\alpha_1)V(G)$ and

$$\alpha_1 V(F_1) + (1-\alpha_1)V(G) = V\big(\alpha_1 F_1 + (1-\alpha_1)G\big).$$

So choose $\pi \in \mathcal{C}(\alpha_1 F_1 + (1-\alpha_1)G)$. Then

$$\big\langle \alpha_1 \varphi_{F_1} + (1-\alpha_1)\varphi_G, \pi\big\rangle - V\big(\alpha_1 F_1 + (1-\alpha_1)G\big) = c(\pi) \geq \langle \varphi_{F_1}, \pi\rangle - V(F_1).$$

Replacing $V(F_1)$ with $(1/\alpha_1)V(\alpha_1 F_1 + (1-\alpha_1)G) - ((1-\alpha_1)/\alpha_1)V(G)$ and, rearranging, we get

$$(1-\alpha_1)\langle \varphi_G, \pi\rangle - \frac{1-\alpha_1}{\alpha_1}V(G) \geq (1-\alpha_1)\langle \varphi_{F_1}, \pi\rangle - \frac{1-\alpha_1}{\alpha_1}V\big(\alpha_1 F_1 + (1-\alpha_1)G\big).$$

Multiplying both sides by $\alpha_1/(1-\alpha_1)$, adding $\langle \varphi_G, \pi\rangle$ to both sides, and rearranging, we get

$$\langle \varphi_G, \pi\rangle - V(G) \geq \big\langle \alpha_1 \varphi_{F_1} + (1-\alpha_1)\varphi_G, \pi\big\rangle - V\big(\alpha_1 F_1 + (1-\alpha_1)G\big),$$

which implies that $\langle \varphi_G, \pi\rangle - V(G) \geq c(\pi)$. Hence, it must be that $\pi \in \mathcal{C}(G)$. The analogous argument shows that $\pi \in \mathcal{C}(F_1)$, so that

$$\mathcal{C}\big(\alpha_1 F_1 + (1-\alpha_1)G\big) \subset \mathcal{C}(F) \cap \mathcal{C}(G).$$

Since $\alpha_2 V(F_2) + \cdots + \alpha_n V(F_n) = (1-\alpha_1)V(G)$, by the inductive assumption, $\mathcal{C}(G) \subset \mathcal{C}(F_i)$ for all $i = 2, \ldots, n$. We conclude that $\mathcal{C}(\alpha_1 F_1 + \cdots + \alpha_n F_n) \subset \mathcal{C}(F_i)$ for all $i = 1, \ldots, n$.

Since $\mathcal{C}(\alpha_1 F_1 + \cdots + \alpha_n F_n)$ is nonempty, (ii) implies (iii).

To see that (iii) implies (i), choose some $\pi \in \mathcal{C}(F_1) \cap \cdots \cap \mathcal{C}(F_n)$. Then

$$\alpha_1 V(F_1) + \cdots + \alpha_n V(F_n) = \langle \alpha_1 \varphi_{F_1} + \cdots + \alpha_n \varphi_{F_n}, \pi\rangle - c(\pi) \leq V(\alpha_1 F_1 + \cdots + \alpha_n F_n).$$

By Jensen's inequality, the convexity of V implies

$$\alpha_1 V(F_1) + \cdots + \alpha_n V(F_n) \geq V(\alpha_1 F_1 + \cdots + \alpha_n F_n)$$

and, therefore, (i) holds. □

Proof of Corollary 1 It is straightforward to prove that a constrained-information preference $\succsim$ satisfies Axioms 1–8, and so we omit this direction of the proof. For the converse, suppose $\succsim$ satisfies Axioms 1–8. By Theorem 1, $\succsim$ is a rationally inattentive preference represented by some $(u, \bar{p}, c)$, and by Theorem 2 it is without loss of generality to assume that c is canonical. We will show that adding Axiom 8 implies that $c(\pi) \in \{0, \infty\}$ for all $\pi \in \Pi(\bar{p})$, so that $\succsim$ is a constrained-information preference.

Let $V : \Phi_{\mathbb{F}} \to \mathbb{R}$ be given by $V(\varphi_F) = \max_{\pi \in \Pi(\bar{p})} \langle \varphi_F, \pi\rangle - c(\pi)$. We first show that Axiom 8 implies that V is homogeneous, that is, $V(\alpha\varphi_F) = \alpha V(\varphi_F)$ for all $F \in \mathbb{F}$ and $\alpha > 0$.

Fix some F such that $\varphi_F \geq 0$. First assume that $\alpha \in [0, 1]$ and let $f \in \mathcal{F}$ be such that $f \sim F$. By Axiom 8, we have that $F \sim \alpha F + (1-\alpha)f$, and so

$$V(\varphi_F) = V\big(\alpha\varphi_F + (1-\alpha)\varphi_f\big) = V(\alpha\varphi_F) + (1-\alpha)V(\varphi_f),$$

where the second equality comes from the translation invariance of V (see Claim 6 in the proof of Theorem 1). Substituting $V(\varphi_f) = V(\varphi_F)$ and rearranging terms, we obtain $V(\alpha\varphi_F) = \alpha V(\varphi_F)$. Now assume that $\alpha > 1$. Then

$$V\big(\alpha^{-1}(\alpha\varphi_F)\big) = \alpha^{-1}V(\alpha\varphi_F) \quad \Rightarrow \quad V(\alpha\varphi_F) = \alpha V(\varphi_F),$$

where the first equality follows from the case $\alpha \in (0,1)$. Hence, $V(\alpha\varphi_F) = \alpha V(\varphi_F)$ for any F such that $\varphi_F \geq 0$ and $\alpha > 0$.

Now we show that if V is homogeneous, then $c(\pi) \in \{0, \infty\}$ for all $\pi \in \Pi(\bar{p})$.

Fix some $\pi \in \Pi(\bar{p})$. Notice that if $c(\pi) > 0$, then there exists a menu G such that $\langle \varphi_G, \pi\rangle - V(\varphi_G) > 0$. By translation invariance of V, we can assume without loss of generality that $\varphi_G \geq 0$. Now notice that

$$\begin{aligned} c(\pi) &= \sup_{F\in\mathbb{F}} \langle \varphi_F, \pi\rangle - V(\varphi_F) \\ &\geq \sup_{\alpha\in(0,\infty)} \langle \alpha\varphi_G, \pi\rangle - V(\alpha\varphi_G) \\ &= \sup_{\alpha\in(0,\infty)} \alpha\big(\langle \varphi_G, \pi\rangle - V(\varphi_G)\big) = \infty. \end{aligned}$$

Finally, notice that if $c(\pi)$ is always equal to 0 or ∞, we can write

$$V(\varphi_F) = \max_{\pi\in\Pi(\bar{p})} \langle \varphi_F, \pi\rangle - c(\pi) = \max_{\{\pi : c(\pi)=0\}} \langle \varphi_F, \pi\rangle$$

so that $\succsim$ is a constrained-information preference.

Proof of Corollary 2 It is straightforward to prove that a subjective-learning preference $\succsim$ satisfies Axioms 1–4, 5b, 6, and 7. For the converse, suppose $\succsim$ satisfies Axioms 1–4, 5', 6, and 7. Clearly Axiom 5b implies Axiom 5, and so by Theorem 1, $\succsim$ is a rationally inattentive preference represented by some $(u, \bar{p}, c)$, and by Theorem 2 it is without loss of generality to assume that c is canonical. We will show that Axiom 5b implies that $\succsim$ is a subjective-learning preference.

By Axiom 5b, for all menus F and G, and $\alpha \in (0,1)$,

$$\alpha V(F) + (1-\alpha)V(G) = V\big(\alpha F + (1-\alpha)G\big).$$

By induction it is easy to see that for all menus $F_1, \ldots, F_n$, and $\alpha_1, \ldots, \alpha_n > 0$ summing to 1,

$$V(\alpha_1 F_1 + \cdots + \alpha_n F_n) = \alpha_1 V(F_1) + \cdots + \alpha_n V(F_n).$$

By Lemma 1, $\mathcal{C}(F_1) \cap \cdots \cap \mathcal{C}(F_n) \neq \varnothing$. Hence, the collection of closed sets $\{\mathcal{C}(F) : F \in \mathbb{F}\}$ has the finite intersection property. Since $\Pi(\bar{p})$ is compact, we conclude that $\bigcap_{F\in\mathbb{F}} \mathcal{C}(F) \neq \varnothing$. Hence, we can choose some $\pi \in \bigcap_{F\in\mathbb{F}} \mathcal{C}(F)$ with $c(\pi) < \infty$, so that for all menus F and G,

$$\langle \varphi_F, \pi\rangle \geq \langle \varphi_G, \pi\rangle \quad \Leftrightarrow \quad V(F) \geq V(G),$$

which implies that $\succsim$ is a subjective-learning preference.

Proof of Theorem 2 Let $(u, \bar{p}, c)$ represent a rationally inattentive preference. First, we verify that, without loss of generality, we can assume that c is canonical. Next—and this is the core of the proof—we show that if c is canonical, then c satisfies (2).

Claim 8. *There exists c' canonical such that $(u, \bar{p}, c')$ represent the same preference as $(u, \bar{p}, c)$.*

Proof. Clearly adding a constant to the cost function c does not affect preferences; therefore without loss of generality, assume $0 \in c(\Pi(\bar{p}))$. Define the value function $V : \mathbb{F} \to \mathbb{R}$ such that

$$V(F) = \max_{\pi \in \Pi(\bar{p})} \langle \varphi_F, \pi \rangle - c(\pi) \quad \forall F \in \mathbb{F}.$$

Choose $c' : \Pi(\bar{p}) \to (-\infty, \infty]$ such that

$$c'(\pi) = \sup_{F \in \mathbb{F}} \langle \varphi_F, \pi \rangle - V(F) \quad \forall F \in \mathbb{F}.$$

Since $0 \in c(\Pi(\bar{p}))$, then $\langle \varphi_f, \pi \rangle = \varphi_f(\bar{p}) = V(f)$ for every $\pi \in \Pi(\bar{p})$ and $f \in \mathcal{F}$. This implies that $c'(\pi) \geq 0$ for every $\pi \in \Pi(\bar{p})$. Moreover, $\langle \varphi_F, \pi_0 \rangle = \max_{f \in F} \varphi_f(\bar{p}) \leq V(F)$ for all $F \in \mathbb{F}$. Therefore, $c'(\pi_0) \leq 0$ and so $c'(\pi_0) = 0$. It is also clear that c' is convex, lower semicontinuous, and Blackwell monotone. Therefore, c' is canonical. Finally, a standard variational argument shows that

$$V(F) = \max_{\pi \in \Pi(\bar{p})} \langle \varphi_F, \pi \rangle - c'(\pi) \quad \forall F \in \mathbb{F},$$

which implies that $(u, \bar{p}, c')$ represents the same preference as $(u, \bar{p}, c)$, completing the proof. □

For the rest of the proof, assume that c is canonical. With some abuse of notation, define the functional $V : \Phi \to \mathbb{R}$ such that

$$V(\varphi) = \max_{\pi \in \Pi(\bar{p})} \langle \varphi, \pi \rangle - c(\pi) \quad \forall \varphi \in \Phi. \tag{4}$$

Fix $\pi \in \Pi(\bar{p})$. From (4) we see that $c(\pi) \geq \langle \varphi_F, \pi \rangle - V(\varphi_F)$ for each menu F, and so $c(\pi) \geq \sup_{F \in \mathbb{F}} \langle \varphi_F, \pi \rangle - V(\varphi_F)$. To show that $c(\pi) = \sup_{F \in \mathbb{F}} \langle \varphi_F, \pi \rangle - V(\varphi_F)$, we therefore need to show the reverse inequality. It is sufficient to show that for each $c(\pi) > \alpha \geq 0$, we have $\sup_{F \in \mathbb{F}} \langle \varphi_F, \pi \rangle - V(\varphi_F) \geq \alpha$. Since $\Phi_{\mathbb{F}} + \mathbb{R}$ is dense in Φ and V is continuous and translation invariant, it is enough to verify that $\sup_{\varphi \in \Phi} \langle \varphi, \pi \rangle - V(\varphi) \geq \alpha$ for each $c(\pi) > \alpha \geq 0$. To do so, fix $0 \leq \alpha < c(\pi)$, and define the sets

$$\operatorname{epi} c = \big\{(\rho, \beta) \in \Pi(\bar{p}) \times \mathbb{R} : c(\rho) \leq \beta\big\} \quad \text{and} \quad \Phi^*_{\pi,\alpha} = \big\{(\pi + \rho, \alpha) : \rho \in \Phi^*\big\},$$

where $\operatorname{epi} c$ is the epigraph of c. Since c is convex, is lower semicontinuous, and $c(\pi_0) = 0$, $\operatorname{epi} c$ is nonempty, convex, and closed. Since Φ^* is convex and closed, $\Phi^*_{\pi,\alpha}$ is convex and closed. Adapting an argument in Sarver (2014, Proof of Claim 1), we first establish the following claim about the difference between $\operatorname{epi} c$ and $\Phi^*_{\pi,\alpha}$.

Claim 9. *The origin* $(0,0) \in \mathrm{ca}(\Delta(\Omega)) \times \mathbb{R}$ *is not a limit point of* $\mathrm{epi}\, c - \Phi^*_{\pi,\alpha}$.

Proof. Define the sets

$$A = \mathrm{epi}\, c \cap \big(\Pi(\bar{p}) \times [0, \alpha+1]\big) \quad \text{and} \quad B = \mathrm{epi}\, c \cap \big(\Pi(\bar{p}) \times [\alpha+1, \infty)\big).$$

Since $c \geq 0$, we have that $\mathrm{epi}\, c = A \cup B$, so that $\mathrm{epi}\, c - \Phi^*_{\pi,\alpha}$ is covered by $A - \Phi^*_{\pi,\alpha}$ and $B - \Phi^*_{\pi,\alpha}$. It is clear that $B - \Phi^*_{\pi,\alpha}$ is a subset of $\mathrm{ca}(\Delta(\Omega)) \times [1, \infty)$, which is a closed set that does not contain $(0,0)$, and so $(0,0)$ is not a limit point of $B - \Phi^*_{\pi,\alpha}$. Hence, it is sufficient to show that $(0,0)$ is not a limit point of $A - \Phi^*_{\pi,\alpha}$.

Observe that A is the intersection of the closed set $\mathrm{epi}\, c$ and the compact set $\Pi(\bar{p}) \times [0, \alpha+1]$: hence A is compact. Since $A - \Phi^*_{\pi,\alpha}$ is the difference of a compact set and a closed set, $A - \Phi^*_{\pi,\alpha}$ is closed. So it is sufficient to show that $(0,0)$ does not belong to $A - \Phi^*_{\pi,\alpha}$ or, equivalently, that A and $\Phi^*_{\pi,\alpha}$ are disjoint. Pick $\rho \in \Phi^*$ such that $\pi + \rho \in \Pi(\bar{p})$. Since $\langle \varphi, \rho \rangle \geq 0$ for each $\varphi \in \Phi$, then $\pi + \rho \trianglerighteq \pi$. Since c is monotone in the Blackwell order and $c(\pi) > \alpha$, $(\pi + \rho, \alpha)$ does not belong to the epigraph of c, which implies that $(\pi + \rho, \alpha)$ does not belong to A. Hence, A and $\Phi^*_{\pi,\alpha}$ are disjoint, and so $(0,0)$ is not a limit point of $A - \Phi^*_{\pi,\alpha}$. □

Since $(0,0)$ is not a limit point of $\mathrm{epi}\, c - \Phi^*_{\pi,\alpha}$ and $\mathrm{epi}\, c - \Phi^*_{\pi,\alpha}$ is a nonempty convex set (being the difference of two nonempty convex sets), a separating hyperplane theorem (Aliprantis and Border 2006, Theorem 5.79) guarantees the existence of a function $\varphi \in C(\Delta(\Omega))$, and real numbers γ and κ such that

$$\langle \varphi, \rho_1 \rangle - \langle \varphi, \pi + \rho_2 \rangle + \gamma(\beta - \alpha) \leq \kappa < \langle \varphi, 0 \rangle + \gamma(0) = 0 \tag{5}$$

for each $(\rho_1, \beta) \in \mathrm{epi}\, c$ and each $\rho_2 \in \Phi^*$. In what follows, $\mathrm{dom}\, c$ denotes the (nonempty) effective domain of c, that is, $\mathrm{dom}\, c = \{\pi \in \Pi(\bar{p}) : c(\pi) < \infty\}$.

Claim 10. *In expression (5),* $\gamma \leq 0$ *and* φ *is convex.*

Proof. Suppose, for contradiction, that $\gamma > 0$. Then we can take $\rho_2 = 0$, fix some $\rho_1 \in \mathrm{dom}\, c$, let β go to infinity, and contradict (5).

Again, for contradiction, suppose that $\langle \varphi, \rho_2 \rangle < 0$ for some $\rho_2 \in \Phi^*$. Since Φ^* is a cone, $n\rho_2 \in \Phi^*$ for each natural number n. Fix any $(\rho_1, \beta) \in \mathrm{epi}\, c$ and observe that for n large enough, we must have

$$\langle \varphi, \rho_1 \rangle - \langle \varphi, \pi \rangle + \gamma(\beta - \alpha) > \kappa + n\langle \varphi, \rho_2 \rangle,$$

contradicting (5). Since $\langle \varphi, \rho_2 \rangle \geq 0$ for each $\rho_2 \in \Phi^*$ and $\Phi = \Phi^{**}$, we conclude that φ is convex (see Appendix A.1). □

We conclude the proof by showing that expression (5) implies $\sup_{\varphi \in \Phi} \langle \varphi, \pi \rangle - V(\varphi) \geq \alpha$. By Claim 10, we can focus on the cases $\gamma < 0$ and $\gamma = 0$, which are treated separately in the following two claims.

Claim 11. *Suppose that* $\gamma < 0$ *in expression (5). Then* $\sup_{\varphi \in \Phi} \langle \varphi, \pi \rangle - V(\varphi) \geq \alpha$.

PROOF. Define $\psi = -\varphi/\gamma$. By Claim 10, φ is convex and $-\gamma > 0$ by hypothesis: hence $\psi \in \Phi$. From expression (5),

$$\langle \psi, \rho_1 \rangle - \langle \psi, \pi + \rho_2 \rangle - (\beta - \alpha) \leq 0$$

for each $(\rho_1, \beta) \in \operatorname{epi} c$ and each $\rho_2 \in \Phi^*$. Taking $\beta = c(\rho_1)$ and $\rho_2 = 0$,

$$\langle \psi, \rho_1 \rangle - c(\rho_1) \leq \langle \psi, \pi \rangle - \alpha \quad \forall \rho_1 \in \operatorname{dom} c \quad \Rightarrow \quad V(\psi) \leq \langle \psi, \pi \rangle - \alpha.$$

Hence we conclude that $\sup_{\varphi \in \Phi} \langle \varphi, \pi \rangle - V(\varphi) \geq \alpha$. □

CLAIM 12. *Suppose that $\gamma = 0$ in (5). Then $\sup_{\psi \in \Phi} \langle \psi, \pi \rangle - V(\psi) \geq \alpha$.*

PROOF. Substituting $\rho_2 = 0$ in (5), we obtain $\langle \varphi, \rho_1 \rangle \leq \kappa + \langle \varphi, \pi \rangle$ for each $\rho_1 \in \operatorname{dom} c$, which implies $\inf_{\rho_1 \in \operatorname{dom} c} \langle \varphi, \pi - \rho_1 \rangle > 0$.

Since $c(\pi_0) = 0$, it follows that for each natural number n,

$$\langle n\varphi, \pi \rangle - V(n\varphi) = \langle n\varphi, \pi \rangle - \Big(\sup_{\rho_1 \in \operatorname{dom} c} \langle n\varphi, \rho_1 \rangle - c(\rho_1) \Big) \geq n \inf_{\rho_1 \in \operatorname{dom} c} \langle \varphi, \pi - \rho_1 \rangle.$$

It therefore follows from $\inf_{\rho_1 \in \operatorname{dom} c} \langle \varphi, \pi - \rho_1 \rangle > 0$ that $\sup_{\psi \in \Phi} \langle \psi, \pi \rangle - V(\psi) = \infty \geq \alpha$. □

Proof of Corollary 3 Assume that the rationally inattentive preference $\succsim$ is represented both by $(u, \bar{p}, c)$ and $(u', \bar{p}', c')$, where c and c' are canonical. Since the restriction of $\succsim$ to acts has an expected utility representation (Lemma 1), it follows that $\bar{p} = \bar{p}'$ and there exist some $\alpha > 0$ and $\beta \in \mathbb{R}$ such that $u' = \alpha u + \beta$. By Theorem 2, for all $\pi \in \Pi(\bar{p})$,

$$\begin{aligned} c(\pi) &= \sup_{F \in \mathbb{F}} \Big[\Big\langle \max_{f \in F} \langle u(f), p \rangle, \pi \Big\rangle - u(x_F) \Big] \\ &= \sup_{\mathbb{F}} \Big[\Big\langle \max_{f \in F} \langle \alpha u'(f) + \beta, p \rangle, \pi \Big\rangle - \alpha u'(x_F) - \beta \Big] \\ &= \alpha \sup_{\mathbb{F}} \Big[\Big\langle \max_{f \in F} \langle u'(f), p \rangle, \pi \Big\rangle - u'(x_F) \Big] \\ &= \alpha c'(\pi). \end{aligned}$$

Proof of Corollary 4 It is straightforward to prove that (ii) implies $(u_1, \bar{p}_1) \approx (u_2, \bar{p}_2)$ so we can normalize $(u_1, \bar{p}_1) = (u_2, \bar{p}_2)$ without loss of generality. For $i = 1, 2$, let $V_i : \mathbb{F} \to \mathbb{R}$ represent $(u_i, \bar{p}_i, c_i)$ in the sense of Definition 2. From Theorem 2, it is clear that $V_1 \geq V_2$ if and only if $c_1 \leq c_2$. So it is enough to show that (i) is equivalent to $V_1 \geq V_2$.

Assume first that (i) holds. For all menus F, choose h so that $V_2(F) = V_2(h)$. Then it has to be the case that $V_1(F) \geq V_1(h) = V_2(F)$; hence $V_1 \geq V_2$.

Alternatively, suppose that $V_1 \geq V_2$. Choose a menu F and an act h: whenever $V_2(F) \geq V_2(h)$, since $V_1(F) \geq V_2(F)$ and $V_2(g) = V_1(g)$, we have that $V_1(F) \geq V_1(g)$, so (i) holds.

Lemma 2 The following lemma, which is used in the proof of Theorem 3, provides an alternative characterization of the Blackwell order over sets in Definition 4.

LEMMA 2. *Let* Π_1 *and* Π_2 *be two subsets of* $\Pi(\bar{p})$*, where* Π_1 *is convex and compact. Then* $\Pi_1 \trianglerighteq \Pi_2$ *if and only if, for all* $\pi_2 \in \Pi_2$*, there exists* $\pi_1 \in \Pi_1$ *such that* π_1 *is Blackwell more informative than* π_2*.*

PROOF. We show the contrapositives.

First suppose that $\sup_{\pi_1 \in \Pi_1} \langle \varphi, \pi_1 \rangle < \sup_{\pi_2 \in \Pi_2} \langle \varphi, \pi_2 \rangle$ for some $\varphi \in \Phi$. Since Π_1 is compact, $\sup_{\pi_1 \in \Pi_1} \langle \varphi, \pi_1 \rangle$ is attained. Hence, there is $\pi_2 \in \Pi_2$ such that $\langle \varphi, \pi_2 \rangle > \langle \varphi, \pi_1 \rangle$ for all $\pi_1 \in \Pi_1$.

Now suppose that there is some $\pi_2 \in \Pi_2$ such that there is no $\pi_1 \in \Pi_1$ with $\pi_1 \trianglerighteq \pi_2$. This means that Π_1 is disjoint from $\pi_2 + \Phi^*$. Since Π_1 is convex and compact and $\pi_2 + \Phi^*$ is convex and closed, there exists a hyperplane strongly separating these two sets (Aliprantis and Border 2006, Theorem 5.79). This means that there is a nonzero $\varphi \in C(\Delta(\Omega))$ such that, for every $\pi \in \Phi^*$ and $\pi_1 \in \Pi_1$, $\langle \varphi, \pi_2 + \pi \rangle > \langle \varphi, \pi_1 \rangle$. In particular, $\langle \varphi, \pi \rangle > \langle \varphi, \pi_1 - \pi_2 \rangle$ for every $\pi \in \Phi^*$. Since Φ^* is a cone, it must be that $\langle \varphi, \alpha\pi \rangle > \langle \varphi, \pi_1 - \pi_2 \rangle$ for every $\alpha > 0$ and for every $\pi \in \Phi^*$, which implies that $\langle \varphi, \pi \rangle \geq 0$ for every $\pi \in \Phi^*$. Since $\Phi^{**} = \Phi$, we have that $\varphi \in \Phi$ (see Appendix A.1). Since $0 \in \Phi^*$, it follows that $\langle \varphi, \pi_2 \rangle > \langle \varphi, \pi_1 \rangle$ for every $\pi_1 \in \Pi_1$, and so $\sup_{\pi_1 \in \Pi_1} \langle \varphi, \pi_1 \rangle < \sup_{\pi_2 \in \Pi_2} \langle \varphi, \pi_2 \rangle$. □

Proof of Theorem 3 To prove that (i) implies (ii), assume that DM1 acquires more information than DM2 (in particular, $(u_1, \bar{p}_1) = (u_2, \bar{p}_2)$). Fix a pair of menus F and G, an act f, and $\alpha \in [0, 1]$. Define, for $i = 1, 2$, the function $W_i(\epsilon) = V_i(\alpha F + (1-\alpha)(\epsilon G + (1-\epsilon)f))$ for all $\epsilon \in [0, 1]$, and observe that

$$W_i(\epsilon) = \max_{\pi \in \Pi(\bar{p})} \langle \alpha\varphi_F, \pi \rangle + \big\langle (1-\alpha)\varphi_f, \pi \big\rangle + \epsilon \big\langle (1-\alpha)(\varphi_G - \varphi_f), \pi \big\rangle - c_i(\pi).$$

By an envelope theorem (Milgrom and Segal 2002, Theorem 2),

$$W_i(1) - W_i(0) = \int_0^1 \big\langle (1-\alpha)(\varphi_G - \varphi_f), \pi_i(\epsilon) \big\rangle \, d\epsilon,$$

where $\epsilon \mapsto \pi_i(\epsilon)$ is any function mapping $[0, 1]$ into $\Pi(\bar{p})$ that satisfies

$$\pi_i(\epsilon) \in \partial V_i\big(\alpha F + (1-\alpha)\big(\epsilon G + (1-\epsilon)f\big)\big) \quad \forall \epsilon \in [0, 1].$$

Since DM1 acquires more information than DM2 and $(1-\alpha)(\varphi_G - \varphi_f) \in \Phi$, for all ϵ we can choose $\pi_1(\epsilon)$ and $\pi_2(\epsilon)$ so that

$$\big\langle (1-\alpha)(\varphi_G - \varphi_f), \pi_1(\epsilon) \big\rangle \geq \big\langle (1-\alpha)(\varphi_G - \varphi_f), \pi_2(\epsilon) \big\rangle$$

by Lemma 2. Hence, by monotonicity of the integral,

$$\int_0^1 \big\langle (1-\alpha)(\varphi_G - \varphi_f), \pi_1(\epsilon) \big\rangle \, d\epsilon \geq \int_0^1 \big\langle (1-\alpha)(\varphi_G - \varphi_f), \pi_2(\epsilon) \big\rangle \, d\epsilon.$$

Therefore,

$$V_1\big(\alpha F + (1-\alpha)G\big) - V_1\big(\alpha F + (1-\alpha)f\big) \geq V_2\big(\alpha F + (1-\alpha)G\big) - V_2\big(\alpha F + (1-\alpha)f\big).$$

Now, to prove that (ii) implies (i), assume that for each pair of menus F and G, for each act f, and for each $\alpha \in [0,1]$,

$$\alpha F + (1-\alpha)G \succsim_2 \alpha F + (1-\alpha)f \quad \Rightarrow \quad \alpha F + (1-\alpha)G \succsim_1 \alpha F + (1-\alpha)f.$$

Observe that this condition implies that DM1 has a stronger desire for flexibility than DM2 (in the sense of Corollary 4(ii)) . Therefore, without loss of generality $(u_1, \bar{p}_1) = (u_2, \bar{p}_2)$.

For $i = 1, 2$, define the functional $V_i : C(\Delta(\Omega)) \to \mathbb{R}$ such that

$$V_i(\varphi) = \max_{\pi \in \Pi(\bar{p})} \langle \varphi, \pi \rangle - c_i(\pi) \quad \forall \varphi \in C\big(\Delta(\Omega)\big).$$

The functional V_i is a niveloid, hence is continuous. Moreover, since c_i is canonical, $\mathcal{C}_i(F)$ coincides with the subdifferential, $\partial V_i(\varphi_F)$, of V_i at φ_F.

Claim 13. *For each pair $\varphi, \tilde{\varphi} \in \Phi$, for each $\psi \in \Phi_{\mathcal{F}} + \mathbb{R}$, and for each $\alpha \in [0,1]$,*

$$V_2\big(\alpha\varphi + (1-\alpha)\tilde{\varphi}\big) \geq V_2\big(\alpha\varphi + (1-\alpha)\psi\big) \quad \Rightarrow \quad V_1\big(\alpha\varphi + (1-\alpha)\tilde{\varphi}\big) \geq V_1\big(\alpha\varphi + (1-\alpha)\psi\big).$$

Proof. First assume that for some menus F and G, act f, and real numbers β, γ, and δ we have

$$\varphi = \varphi_F + \beta, \qquad \tilde{\varphi} = \varphi_G + \gamma, \quad \text{and} \quad \psi = \varphi_f + \delta.$$

Choose $\epsilon \in \mathbb{R}$ large enough so that $\beta + \epsilon, \gamma + \epsilon, \delta + \epsilon \geq 0$. Then it follows that

$$\varphi + \epsilon, \qquad \tilde{\varphi} + \epsilon \in \Phi_{\mathbb{F}} \quad \text{and} \quad \psi + \epsilon \in \Phi_{\mathcal{F}}.$$

Therefore,

$$\begin{aligned}
&V_2\big(\alpha\varphi + (1-\alpha)\tilde{\varphi}\big) \geq V_2\big(\alpha\varphi + (1-\alpha)\psi\big) \\
&\quad \Rightarrow \quad V_2\big(\alpha(\varphi+\epsilon) + (1-\alpha)(\tilde{\varphi}+\epsilon)\big) \geq V_2\big(\alpha(\varphi+\epsilon) + (1-\alpha)(\psi+\epsilon)\big) \\
&\quad \Rightarrow \quad V_1\big(\alpha(\varphi+\epsilon) + (1-\alpha)(\tilde{\varphi}+\epsilon)\big) \geq V_1\big(\alpha(\varphi+\epsilon) + (1-\alpha)(\psi+\epsilon)\big) \\
&\quad \Rightarrow \quad V_1\big(\alpha\varphi + (1-\alpha)\tilde{\varphi}\big) \geq V_1\big(\alpha\varphi + (1-\alpha)\psi\big).
\end{aligned}$$

Now assume $\varphi, \tilde{\varphi} \in \Phi$, and observe that $V_2(\alpha\varphi + (1-\alpha)\tilde{\varphi}) \geq V_2(\alpha\varphi + (1-\alpha)\psi)$ implies that for each $\eta > 0$,

$$V_2\big(\alpha\varphi + (1-\alpha)\tilde{\varphi}\big) > V_2\big(\alpha\varphi + (1-\alpha)\psi\big) - (1-\alpha)\eta.$$

Alternatively, if, for each $\eta > 0$, we have $V_1(\alpha\varphi + (1-\alpha)\tilde{\varphi}) \geq V_1(\alpha\varphi + (1-\alpha)\psi) - (1-\alpha)\eta$, then it follows that

$$V_1\big(\alpha\varphi + (1-\alpha)\tilde{\varphi}\big) \geq V_1\big(\alpha\varphi + (1-\alpha)\psi\big).$$

So choose $\eta > 0$ and assume that $V_2(\alpha\varphi + (1-\alpha)\tilde{\varphi}) > V_2(\alpha\varphi + (1-\alpha)(\psi - \eta))$. Observe that $\psi - \eta \in \Phi_{\mathcal{F}} + \mathbb{R}$. Choose sequences $\{\varphi_n\}$ and $\{\tilde{\varphi}_n\}$ in $\Phi_{\mathbb{F}} + \mathbb{R}$ converging to φ and $\tilde{\varphi}$, respectively. By continuity of V_2, eventually

$$V_2\big(\alpha\varphi_n + (1-\alpha)\tilde{\varphi}_n\big) \geq V_2\big(\alpha\varphi_n + (1-\alpha)(\psi - \eta)\big).$$

This implies that eventually $V_1(\alpha\varphi_n + (1-\alpha)\tilde{\varphi}_n) \geq V_1(\alpha\varphi_n + (1-\alpha)(\psi - \eta))$. By continuity of V_1, it follows that $V_1(\alpha\varphi + (1-\alpha)\tilde{\varphi}) \geq V_1(\alpha\varphi + (1-\alpha)(\psi - \eta))$. Since the choice of η was arbitrary,

$$\begin{aligned} &V_2\big(\alpha\varphi + (1-\alpha)\tilde{\varphi}\big) \geq V_2\big(\alpha\varphi + (1-\alpha)\psi\big) \\ &\quad\Rightarrow\quad V_1\big(\alpha\varphi + (1-\alpha)\tilde{\varphi}\big) \geq V_1\big(\alpha\varphi + (1-\alpha)\psi\big). \end{aligned}$$

□

Claim 14. *For all for all $\varphi, \psi \in \Phi$, $V_1(\varphi + \psi) - V_1(\varphi) \geq V_2(\varphi + \psi) - V_2(\varphi)$.*

Proof. For contradiction, suppose there is a real number α such that

$$V_2(\varphi + \psi) - V_2(\varphi) \geq \alpha, \qquad V_1(\varphi + \psi) - V_1(\varphi) < \alpha.$$

Using translation invariance we can rewrite the above statement as

$$\begin{aligned} V_2\left(\frac{1}{2}(2\varphi) + \frac{1}{2}(2\psi)\right) &\geq V_2\left(\frac{1}{2}(2\varphi) + \frac{1}{2}(2\alpha)\right), \\ V_1\left(\frac{1}{2}(2\varphi) + \frac{1}{2}(2\psi)\right) &< V_1\left(\frac{1}{2}(2\varphi) + \frac{1}{2}(2\alpha)\right), \end{aligned}$$

which contradicts Claim 13. □

Now consider menu F and $\varphi \in \Phi$. Exploiting the relation between directional derivatives and subdifferentials (Rockafellar 1974, Theorem 11), we obtain

$$\inf_{\epsilon > 0} \frac{V_i(\varphi_F + \epsilon\varphi) - V_i(\varphi_F)}{\epsilon} = \max_{\pi_i \in \partial V_i(\varphi_F)} \langle \varphi, \pi_i \rangle.$$

Moreover, from Claim 14,

$$\frac{V_1(\varphi_F + \epsilon\varphi) - V_1(\varphi_F)}{\epsilon} \geq \frac{V_2(\varphi_F + \epsilon\varphi) - V_2(\varphi_F)}{\epsilon} \quad \forall \epsilon > 0.$$

Hence,

$$\max_{\pi_1 \in \partial V_1(\varphi_F)} \langle \varphi, \pi_1 \rangle \geq \max_{\pi_2 \in \partial V_2(\varphi_F)} \langle \varphi, \pi_2 \rangle.$$

Since $\varphi \in \Phi$ was arbitrary, it follows that $\partial V_1(\varphi_F) \supseteq \partial V_2(\varphi_F)$.

A.3 *Infinite state space*

Finally, we briefly discuss how our characterization of rationally inattentive preferences can be generalized to an infinite state space.

For a measurable space E, denote by $B(E)$ the set of all real-valued bounded measurable functions defined on E. The real linear space $B(E)$ is normed by the uniform norm. Denote by $\Delta(E)$ the set of all finitely additive probabilities over E. The set $\Delta(E)$ is endowed with the weak topology generated by the family of functions

$$\left\{\mu \mapsto \int_E \varphi(t)\mu(dt) : \varphi \in B(E)\right\}.$$

The topological space $\Delta(E)$ is then endowed with the Borel σ-algebra.

Adapting our previous notation, let Ω be an arbitrary measurable space of states of the world and let X be a convex set of outcomes. Denote by $\mathcal{F}$ the set of all *simple* acts: functions $f : \Omega \to X$ with finite range, such that $f^{-1}(x)$ is measurable for each $x \in X$. Denote by $\mathbb{F}$ the set of all menus: nonempty, finite subsets of $\mathcal{F}$. Definition 2 and Axioms 1–7 then apply exactly as they are formulated in Sections 2 and 3. The following theorem generalizes Theorem 1 to this framework:

Theorem 4. *A binary relation over menus is a rationally inattentive preference if and only if it satisfies Axioms 1–7. Moreover, the information cost function can be chosen to be the canonical one defined as in (2) in Theorem 2.*

There are no conceptual differences between Theorem 4 and Theorem 1. The main technical difference is that the probabilities involved may not be countably additive. To prove Theorem 4, the steps in the proof of Theorem 1 can be followed verbatim, with two exceptions. In Claim 6, we embed convex combinations of finitely many acts in a Euclidean space using the fact that the smallest σ-algebra for which they are measurable has finitely many atoms. In Claim 7, we use the duality between $B(\Delta(\Omega))$ and $ba(\Delta(\Omega))$, the set of signed charges of bounded variation on $\Delta(\Omega)$ (Aliprantis and Border 2006, p. 396).

References

Aliprantis, Charalambos D. and Kim C. Border (2006), *Infinite Dimensional Analysis*. Springer, Berlin. [638, 647, 649, 652]

Blackwell, David (1951), "Comparison of experiments." In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability* (Jerzy Neyman, ed.), 93–102, University of California Press, Berkeley, California. [625]

Blackwell, David (1953), "Equivalent comparisons of experiments." *Annals of Mathematical Statistics*, 24, 265–272. [625]

Caplin, Andrew and Mark Dean (2014), "Revealed preference, rational inattention, and costly information acquisition." National Bureau of Economic Research Working Paper 19876. [623, 633]

Caplin, Andrew and Mark Dean (2015), "Revealed preference, rational inattention, and costly information acquisition." *American Economic Review*, 105, 2183–2203. [623, 635]

Cerreia-Vioglio, Simone, Fabio Maccheroni, Massimo Marinacci, and Aldo Rustichini (2014), "Niveloids and their extensions: Risk measures on small domains." *Journal of Mathematical Analysis and Applications*, 413, 343–360. [638]

Cheremukhin, Anton, Anna Popova, and Antonella Tutino (2015), "A theory of discrete choice with information costs." *Journal of Economic Behavior and Organization*, 113, 34–50. [623, 633]

Cover, Thomas M. and Joy A. Thomas (2006), *Elements of Information Theory*, second edition. Wiley-Interscience, Hoboken, New Jersey. [627, 632]

Dekel, Eddie, Barton L. Lipman, and Aldo Rustichini (2001), "Representing preferences with a unique subjective state space." *Econometrica*, 69, 891–934. [623, 629, 633, 635]

Dillenberger, David, Juan Sebastián Lleras, Philipp Sadowski, and Norio Takeoka (2014), "A theory of subjective learning." *Journal of Economic Theory*, 153, 287–312. [623, 627, 629, 631, 635]

Ellis, Andrew (2016), "Foundations for optimal inattention." Unpublished, The London School of Economics and Political Science. [623]

Ergin, Haluk and Todd Sarver (2015), "Hidden actions and preferences for timing of resolution of uncertainty." *Theoretical Economics*, 10, 489–541. [628]

Ergin, Haluk and Todd Saver (2010), "A unique costly contemplation representation." *Econometrica*, 78, 1285–1339. [623, 625, 628, 629, 630, 633, 635]

Gabaix, Xavier, David Laibson, Guillermo Moloche, and Stephen Weinberg (2006), "Costly information acquisition: Experimental analysis of a boundedly rational model." *American Economic Review*, 96, 1043–1068. [623, 633]

Kreps, David M. (1979), "A representation theorem for 'preference for flexibility'." *Econometrica*, 47, 565–577. [628, 629]

Luo, Yulei (2010), "Rational inattention, long-run consumption risk, and portfolio choice." *Review of Economic Dynamics*, 13, 843–860. [623]

Maccheroni, Fabio, Massimo Marinacci, and Aldo Rustichini (2006), "Ambiguity aversion, robustness, and the variational representation of preferences." *Econometrica*, 74, 1447–1498. [623, 627, 628, 629, 630, 633, 640, 641]

Manzini, Paola and Marco Mariotti (2014), "Imperfect attention and menu evaluation." Unpublished, University of St Andrews Discussion Paper 1319. [628]

Masatlioglu, Yusufcan, Daisuke Nakajima, and Erkut Y. Ozbay (2012), "Revealed attention." *American Economic Review*, 102, 2183–2205. [628]

Matejka, Filip and Alisdair McKay (2015), "Rational inattention to discrete choices: A new foundation for the multinomial logit model." *American Economic Review*, 105, 272–298. [623]

Milgrom, Paul R. and Ilya Segal (2002), "Envelope theorems for arbitrary choice sets." *Econometrica*, 70, 583–601. [649]

Ortoleva, Pietro (2013), "The price of flexibility: Towards a theory of thinking aversion." *Journal of Economic Theory*, 148, 903–934. [628]

Rockafellar, R. Tyrrell (1970), *Convex Analysis*. Princeton University Press, Princeton, New Jersey. [640]

Rockafellar, R. Tyrrell (1974), *Conjugate Duality and Optimization*, volume 16 of CBMS-NSF Regional Conference Series in Applied Mathematics. Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania. [642, 651]

Ruszczyński, Andrzej and Alexander Shapiro (2006), "Optimization of convex risk functions." *Mathematics of Operations Research*, 31, 433–452. [642]

Sarver, Todd (2014), "Optimal reference points and anticipation." Unpublished, Duke University. [633, 646]

Schneider, Rolf (1993), *Convex Bodies: The Brunn–Minkowski Theory*. Cambridge University Press, Cambridge. [639]

Sims, Christopher A. (1998), "Stickiness." *Carnegie-Rochester Conference Series on Public Policy*, 49, 317–356. [622, 627, 637]

Sims, Christopher A. (2003), "Implications of rational inattention." *Journal of Monetary Economics*, 50, 665–690. [622, 623, 627, 637]

Tutino, Antonella (2013), "Rationally inattentive consumption choices." *Review of Economic Dynamics*, 16, 421–439. [623]

Veldkamp, Laura L. (2011), *Information Choice in Macroeconomics and Finance*. Princeton University Press, Princeton, New Jersey. [622]

Wiederholt, Mirko (2010), "Rational inattention." In *The New Palgrave Dictionary of Economics* (Steven N. Durlauf and Lawrence E. Blume, eds.). Palgrave Macmillan. [622]

Co-editor Ran Spiegler handled this manuscript.

Manuscript received 6 October, 2015; final version accepted 8 February, 2016; available online 3 March, 2016.