Rationally Inattentive Preferences*

Henrique de Oliveira[†] Tommaso Denti[‡] Maximilian Mihm[§] Kemal Ozbek[¶]

July 25, 2014

Abstract

A rationally inattentive agent processes information by balancing benefits and costs of attention (Sims [1998, 2003]). Predicting the agent's behavior therefore requires a measure of her attention costs, but these costs can be difficult to identify as they incorporate hidden factors - such as time, effort and cognitive resources - that are not directly observable. In this paper, we identify attention costs by characterizing the implications of rational inattention for choices over opportunity sets, where the trade-off between benefits and costs are revealed by attitudes towards flexibility (Kreps [1979]) and temporal resolution of uncertainty (Kreps and Porteus [1978]). Exploiting this connection, we provide a procedure to elicit attention costs using willingness-to-pay data, and indicate how such data could be obtained in dynamic choice environments (e.g., consumption-saving problems) or generated in experimental settings.

Key Words: Blackwell order, flexibility, menu choice, rational inattention, temporal resolution of uncertainty.

*This paper merges and improves upon 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, Aaron Bodoh-Creed, Giulia Brancaccio, Enrico Cantoni, David Easley, Ludovica Gazzé, Peter Klibanoff, Samreen Malik, Jawwad Noor, Efe Ok, Pietro Ortoleva, Luciano Pomatto, Marciano Siniscalchi, Debraj Ray, Mirko Wiederholt and seminar participants at Cornell University, NYU, NYUAD, Northwestern University, MIT, RUD 2012 and University of St Andrews for helpful comments and suggestions.

[†]Northwestern University. Email: henriquedeoliveira@u.northwestern.edu.

[‡]MIT. Email: tdenti@mit.edu.

[§]NYU Abu Dhabi. Email: max.mihm@nyu.edu.

[¶]University of St Andrews. Email: kemal.ozbek@st-andrews.ac.uk.

1 Introduction

Economic agents often forgo decision-making benefits of information, due to limitations on attention. For example, consumers purchase goods without considering sales taxes (Chetty, Looney, and Kroft [2009]) or shipping charges (Brown, Hossain, and Morgan [2010]), and investors make portfolio decisions overlooking financial news (DellaVigna and Pollet [2009]; Hirshleifer et al. [2009]).[1]

Rational inattention (Sims [1998, 2003]) is an influential approach to study implications of limited attention in economic environments. A rationally inattentive agent does not observe the current state of economic fundamentals, but is able to select, study and process information to extract a noisy signal about the state. Intuitively, extracting a more precise signal allows the agent to improve her choices but requires a greater input of attention, and the agent must balance the decision-making benefits of information against the costs of attention.[2] Rational inattention is therefore a model of costly information acquisition, but in contrast with traditional models, the costs of information are not necessarily observable and there are no exogenous restrictions on the type of signals the agent can extract. Instead, limitations on information processing are reflected in subjective attention costs, and the agent optimally determines the nature of signals conveying information about states.[3] As a result, rational inattention can accommodate a wide range of deviations from perfect information processing, and accordingly has been applied to study a variety of empirical phenomena, from price stickiness and business-cycle dynamics (Maćkowiak and Wiederholt [2009, 2011]) to portfolio under-diversification (van Nieuwerburgh and Veldkamp [2010]) and discrete-pricing patterns (Matějka [2012]).[4]

In this paper, we propose a framework to study implications of rational inattention and to elicit attention costs from choice data. Attention costs are the key parameter in rational inattention models, and identifying these costs is essential to make predictions about the behavior of the agent. For example, optimal monetary policy rules may depend on the type and severity of attention limitations faced by economic agents (Sims [2010]; Paciello and Wiederholt [2014]), and the evaluation of different policies therefore requires a measure of such limitations. However, since attention costs can incorporate hidden factors - such as time, effort and cognitive resources - a fundamental challenge is to find a framework in which these costs can be inferred from choice behavior.

Our approach to this problem is based on the following idea. Given a set of available alternatives, i.e. an "opportunity set", a rationally inattentive agent would like to reduce uncertainty to make

[1]See DellaVigna [2009] for further discussion of the empirical literature on attention limitations.

[2]A rationally inattentive agent is fully aware of the choices available to her, and is uncertain only about the state of fundamentals (see, e.g., Sims [2010, p. 12-13]). In general, an agent could be inattentive also to the choices available to her (see Section 4.1 for references).

[3]Further discussion of the relation between rational inattention and traditional models of information acquisition is provided by Sims [2006, Section 1A; 2011, p. 161].

[4]Wiederholt [2010] and Veldkamp [2011] survey recent applications of rational inattention models, and provide additional references.

better choices. However, in order to make a more informed choice, the agent incurs a cost for paying attention to information. The value of an opportunity set therefore depends not only on the set of available choices, but also on the agent's subjective costs of attention. Intuitively, an agent with low costs of attention would be willing to accept opportunity sets with more variability in payoffs across states, as the agent can utilize benefits of information at a low cost. On the other hand, an agent with high costs of attention would prefer opportunity sets with less variability, because these sets will require less attention. This intuition suggests a simple elicitation method: determine the agent's willingness-to-pay for various opportunity sets and use this data to reveal her attention costs.

Our analysis formalizes this idea in a framework of preferences over opportunity sets. We take the perspective of an agent when "today" she chooses the opportunity set for "tomorrow", with the prospect of processing information before selecting an alternative. Figure 1 illustrates the order of events we have in mind.

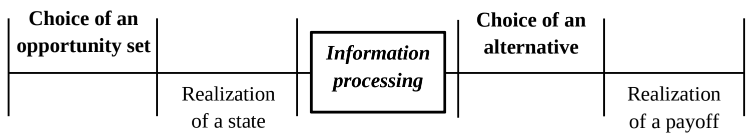

Figure 1: Timeline.

Today, the agent chooses an opportunity set, being uncertain about the future state of fundamentals. Tomorrow, a state realizes, and information becomes available which the agent can process.[5] On the basis of information acquired, the agent can then choose an alternative from her opportunity set and obtain a state-dependent payoff.

The choice data corresponding to this framework arises in many dynamic environments, where agents make choices every period that affect both current well-being and future choice opportunities, and our framework can be interpreted as a "snapshot" of such dynamic choice problems. For instance, consider the consumption-saving problem studied by Sims [2003]; Luo [2010]; Tutino [2012]: the agent pays attention to 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 an opportunity set, which is made before attention is paid to future income. We give an example in Section 3.2 to make this connection precise, and show that costs of attention tomorrow induce a motive for precautionary saving today, which can be used to elicit the agent's costs from her saving behavior.

[5]We follow the rational inattention literature in assuming that the agent cannot process "information about shocks that nature has not drawn yet" (Maćkowiak and Wiederholt [2009, p. 775]). In our framework, this means that the agent can process information only after she has chosen an opportunity set. In general, there could also be states realizing "today", to which the agent could pay attention before the choice of a menu. To focus on the effects of future attention costs on current choices, we consider an agent facing uncertainty only about the future state of fundamentals.

Overview of the analysis

In our framework we study the relation between current choices and future attention limitations for a general model of rational inattention. In particular, two key features distinguish rational inattention from other approaches to model imperfect information-processing, where attention may not be allocated in a value-maximizing way.[6] First, the agent optimally chooses a costly signal for her opportunity set; second, she optimally chooses an alternative after observing a signal realization. In our framework, the optimal choice of a signal induces a *desire for early resolution of uncertainty* (Kreps and Porteus [1978]), while the optimal response to signal realizations induces a *desire for flexibility* (Kreps [1979]). We show that, together with some standard conditions, these behavioral traits characterize the implications of rational inattention in our framework, providing a basis for testing the model (Theorem 1).

We then consider identification of model parameters in our framework (Proposition 1). Specifically, we show that attention costs can be uniquely identified whenever they are monotone in the order of Blackwell [1953] and convex, which are intuitive properties satisfied by cost functions used in applications. Moreover, we propose an elicitation procedure that exploits the optimizing behavior of the agent to approximate her attention costs using "willingness-to-pay" data – which could be collected empirically or generated in experimental settings – and show that, with sufficient data, the procedure elicits the agent's attention costs (Theorem 2).

Given that parameters can be uniquely identified, we can also study comparative statics in our framework. Specifically, we show that comparisons of attention costs and the "tendency" to reallocate attention are revealed, respectively, by attitudes towards flexibility and temporal resolution of uncertainty (Propositions 2-3). Moreover, an overall increase in "attentiveness" is revealed by a trade-off between these two key behaviors (Proposition 4).

Finally, we compare different special cases of the general model: (i) a model in which attention is constrained rather than costly (as, for example, in Sims [2003]), (ii) a model in which the agent faces a binary attention allocation problem, where she can decide whether to acquire an informative signal, at a fixed cost, or remain uninformed (as, for example, in Grossman and Stiglitz [1980]), and (iii) a model in which the agent receives an exogenous signal, and therefore cannot vary the allocation of attention with her opportunity set (as, for example, in Paciello and Wiederholt [2014]). Using our framework, we show how each of these restrictions on the general model are revealed by the agent's attitudes toward temporal resolution of uncertainty (Propositions 5-7).

The paper is organized as follows. We present our formal approach to study rational inattention in Section 2, and provide some illustrative examples in Section 3. Section 4 contains the main results. Our formal analysis extends on recent developments in decision theory, in particular

[6] Early surveys of alternative models of attention limitation are provided by Lipman [1995] and Conlisk [1996]. Gabaix and Laibson [2006] provides a recent application.

Maccheroni, Marinacci, and Rustichini [2006] and Ergin and Sarver [2010], and we discuss related literature in Section 5. Section 6 concludes. Proofs are given in a separate Appendix.

2 Preliminaries

In this section we present our approach to study rational inattention. We first introduce the formal framework, then present a general model of rational inattention, and finally provide a formal definition of rationally inattentive preferences.

2.1 Framework

There is a finite set Ω of *states*, and a convex set X of possible *outcomes*.[7] An *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. 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 abuse of notation, we identify a singleton menu $\{f\}$ with the act $f \in \mathcal{F}$, and a constant act f such that $f(\omega) = x$ for all $\omega \in \Omega$ with the outcome $x \in X$. Moreover, for any menu F, we call an outcome $x_F \in X$ a *certainty equivalent* of F when $x_F \sim F$, and say that a function $V : \mathbb{F} \to \mathbb{R}$ *represents* $\succsim$ if, for all menus F and G, $F \succsim G$ if and only if $V(F) \geq V(G)$.

To interpret the DM's preferences, we have in mind the following order of events. Today, there is uncertainty about the future state of fundamentals. Tomorrow, a state realizes and information becomes available that could be used to reduce (or even resolve) the uncertainty. We take the perspective of the DM when she chooses today her menu (opportunity set) for tomorrow, 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".

Finally, using the convexity of X, we define mixed acts and menus. For $\alpha \in [0,1]$ and acts f and g, we denote by $\alpha f + (1-\alpha)g$ the *mixed act* h such that

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

For $\alpha \in [0,1]$ and menus F and G, we denote by $\alpha F + (1-\alpha)G$ the *mixed menu* H such that

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

[7]We assume that Ω is finite to simplify the exposition. An arbitrary measurable space can be accommodated with some technical modifications.

To provide intuition in our discussions, we often interpret mixed menus in terms of the Anscombe and Aumann [1963] setting, where outcomes $x \in X$ are lotteries on a set of prizes. In terms of our timeline such lotteries represent an additional source of uncertainty, that realizes *after* the DM has chosen an act from the menu. In this interpretation, the mixed menu $\alpha F + (1-\alpha)G$ denotes the opportunity set in which the DM can choose an act $f \in F$ and an act $g \in G$, not knowing yet which act will actually determine her outcome (with probability α it will be f, and with probability $(1-\alpha)$ it will be g).

2.2 Rational inattention

We now formalize how a rationally inattentive DM allocates attention in terms of a general model of rational inattention.

Before a state realizes, the DM's has a prior $\bar{p} \in \Delta(\Omega)$ representing her initial beliefs about the state. After a state realizes, the DM pays attention to extract a noisy signal. Each possible realization of the signal conveys information about the state and induces a posterior belief $p \in \Delta(\Omega)$ from her prior $\bar{p}$. 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 (see Gollier [2004, Chapter 25] or Kamenica and Gentzkow [2011, p. 2594]).[8] As a result, the collection of all possible signals is given by the set

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

Given a menu F, extracting a signal allows the DM to make a more informed choice from F, by maximizing 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,

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

A rationally inattentive DM must balance this benefit against the cost of processing information. These costs are measured by a subjective *attention cost function* $c : \Pi(\bar{p}) \to [0, \infty]$, which associates a cost $c(\pi)$ to each signal $\pi \in \Pi(\bar{p})$. As such, the DM chooses a signal to maximize the value of information, and therefore solves the following optimization problem:

$$\max_{\pi \in \Pi(\bar{p})} \left[b_F(\pi) - c(\pi) \right].$$

[8] Alternatively, a signal could be represented as a random variable S correlated with Ω. The connection between the two representations for signals is the following: the distribution of the random variable $\Omega | S$ is a distribution over posteriors, while for every distribution over posteriors π one can find a random variable S such that π is the distribution of $\Omega | S$.

This model generalizes and encompasses the models of rational inattention studied in the literature, for instance the ones based on the information theoretic notion of entropy.

Example 1. Following Sims [1998], it has been conventional to quantify information in rational inattention models in terms of *Shannon information*, which measures the expected reduction in entropy due to realizations of signal π:

$$I(\pi) = \int_{\Delta(\Omega)} \left(\int_{\Omega} \log \left(\frac{p(\omega)}{\bar{p}(\omega)} \right) \bar{p}(d\omega) \right) \pi(dp),$$

Using I to measure information, attention costs can be specified by $c(\pi) = \lambda I(\pi)$, where $\lambda \geq 0$ is a scale parameter measuring the unit costs of attention (see, e.g., Matejka and McKay [2012]; Yang [2013]).

2.3 Rationally inattentive preferences

We are now ready to formally connect rational inattention with our menu-choice framework, by specifying how a rationally inattentive DM ranks menus.

Definition 1. 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

$$V(F) = \max_{\pi \in \Pi(\bar{p})} \left[\int_{\Delta(\Omega)} \max_{f \in F} \left(\int_{\Omega} u(f(\omega))\, p(d\omega) \right) \pi(dp) - c(\pi) \right] \qquad \forall\, F \in \mathbb{F},$$

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 attention cost function.[9]

Rationally inattentive preferences will be the focus of our analysis. The properties of c and u are standard in the literature. In particular, the properness and lower-semicontinuity of c ensure that the maximization over costly signals is well-defined; the affinity of u permits the use of convex analysis and, when X is a set of lotteries over prizes, corresponds to the assumption of von Neumann-Morgenstern utility; and finally, the unboundedness of u ensures that the benefit of information is not bounded, so that attention costs can be identified also for high-cost signals.

3 Examples

Before proceeding with the formal analysis of rationally inattentive preferences, we give two examples to clarify the framework. The first example provides a simple illustration of the notation

[9]The function u is said to be *unbounded* if its range is either unbounded above or below (not necessarily both), and *affine* if $u(\alpha x + (1-\alpha)y) = \alpha u(x) + (1-\alpha)u(y)$ for all $x, y \in X$ and $\alpha \in [0,1]$. Lower semi-continuity of c is defined in terms of the weak* topology, and the function is said to be proper if $c \neq \infty$.

in Section 2 (acts, menus and signals). The second example adapts our framework to a specific environment: a simple version of the consumption-saving problem used by Sims [2006] to exemplify rational inattention. In both examples we show how the menu choice approach can be used to identify future attention costs from current behavior, under some parametric assumptions.

3.1 A numerical example

Consider a simple illustration of our framework with two states ($\Omega = \{\omega_1, \omega_2\}$), monetary outcomes ($X = \mathbb{R}_+$), and a risk-neutral DM ($u(x) = x$) with a uniform prior ($\bar{p} = (1/2, 1/2)$). Suppose the DM has to choose an act from a menu $F = \{f_1, f_2\} = \{(28, 4), (4, 20)\}$. Before making her choice, the DM can pay attention to information to extract a signal $\pi \in \Pi(\bar{p})$, which induces posteriors $p_1 = (3/4, 1/4)$ and $p_2 = (1/4, 3/4)$ with equal probability. If she extracts the signal π she incurs an attention cost $c(\pi) > 0$, and otherwise she remains uninformed and incurs no cost.

Using signal π allows the DM to choose act $f_i \in F$ for each realization of posterior p_i ($i = 1, 2$) and as a result the benefit of information from signal π is:

$$b_F(\pi) = \left(\frac{1}{2}\right)\left[\left(\frac{3}{4}\right)28 + \left(\frac{1}{4}\right)4\right] + \left(\frac{1}{2}\right)\left[\left(\frac{1}{4}\right)4 + \left(\frac{3}{4}\right)20\right] = 19.$$

On the other hand, if the DM remains uninformed, she could at best obtain an expected utility of 16 by choosing act $f_1 \in F$, which is optimal under her prior. Thus, for menu F it is optimal to incur the attention cost whenever $c(\pi) \leq 3$. To identify $c(\pi)$ we could therefore try to elicit the value of $x \in [16, 19]$ at which the DM is indifferent between menus F and singleton menu $\{(x, x)\}$. When the DM is indifferent for some $x \in (16, 19]$, we can infer that $c(\pi) = 19 - x$; when she is indifferent for $x = 16$, we can infer that $c(\pi) \geq 3$.

3.2 A consumption-saving example

Kreps [1979, Section 4] motivates the analysis of menu-choice with a consumption-saving problem. To illustrate this connection in our specific framework, we consider a rationally inattentive agent facing a three period problem. For simplicity, we assume that the agent receives income only in period 2, W, with realizations w drawn from a standard normal distribution. Before choosing period 2 consumption, x_2, the agent can extract a signal S (a random variable) conveying information about W, at a cost $c(S)$ reflecting her limitations on attention. Facing no borrowing constraints, the agent then chooses period 1 consumption x_1, a signal S and period 2 consumption x_2 (conditional on the realization s of the signal S) to maximize

$$v(x_1) + E\left[v\left(x_2(S)\right) + v\left(W - x_1 - x_2(S)\right)\right] - c(S),$$

where $v : \mathbb{R} \to \mathbb{R}$ is the agent's instantaneous utility on consumption.

In this problem, each consumption stream $x = (x_1, x_2, x_3)$ can be viewed as an outcome, and each income realization w as a state of the world. A consumption decision for the first two periods (x_1 and x_2) maps each state w to an outcome $x(w) = (x_1, x_2, w - x_1 - x_2)$. As a result, for a given x_1, the subsequent choice of x_2 can be viewed as an act (mapping each possible state w to a particular outcome $x(w)$). A choice of x_1 restricts the set of available acts to be chosen later, and can therefore be interpreted as the choice of a menu. The consumption-saving problem therefore represents a specific instance of our general framework: choice of a menu (x_1), realization of a state (w), allocation of attention (S), choice of an act ($x_2(s)$), and realization of a state-dependent outcome ($x(w)$).

Consumption-saving problems have been studied extensively in the literature on rational inattention (e.g, Sims [2003, 2006]; Luo [2010]; Tutino [2012]), focusing on the effect of current attention costs on current consumption choices (i.e., how c affects x_2). Instead, our approach to study rational inattention focuses on the effect of future attention costs on current consumption (i.e., how c affects x_1), allowing us to highlight implications of rational inattention which have not been studied in the existing literature.

To illustrate, we make a number of simplifying assumptions so that the problem can be solved analytically. First, we restrict the agent to extract signals S that are jointly standard normal with W. Second, we quantify attention costs by the Shannon information $I(S, W)$, scaled by a parameter $\lambda > 0$.[10] With these assumptions, the dependence between S and W is fully described by the correlation coefficient $\rho \in [-1, 1]$, and the attention costs are given as $c(S) = -\lambda \log \sqrt{1 - \rho^2} \in [0, \infty]$. Finally, we assume that the agent's utility is exponential,$v(z) = -e^{-z}$ for all $z \in \mathbb{R}$.[11]

The effect of attention costs on period 2 consumption

We first consider the problem studied in the previous literature: how attention costs affect consumption in period 2, for a given x_1. If the agent observes a signal realization s, her choice of consumption in period 2 is given by

$$x_2(s) = \frac{1}{2}\left((\rho s - x_1) - \frac{1}{2}\left(1 - \rho^2\right)\right).$$

The first component above, $(\rho s - x_1)$, is the expected wealth in period 2 conditional on $S = s$, reflecting the agent's preference for consumption smoothing ($v'' < 0$). The second component, $-\left(1 - \rho^2\right)$, reflects a precautionary saving motive when there is remaining uncertainty about income (when the signal S is not perfectly correlated with W, i.e., $\rho^2 < 1$), since the agent is prudent ($v''' > 0$, Kimball [1990]). The agents choice of an optimal signal S can be identified

[10] See Example 1 in Section 2.2.

[11] Derivations to support the following discussions are given in Appendix A.3.

with its precision ρ^2, which is decreasing in the unit cost of attention λ. The optimal period 2 consumption x_2 is therefore a decreasing function of λ, so that attention costs induce precautionary savings in period 2 (Sims [2006]; Luo [2010]).

The effect of attention costs on period 1 consumption

We can now consider the implications of attention limitations on first period consumption. Incorporating the optimal $x_2(s)$, the agent chooses x_1 and ρ^2 to maximize

$$v(x_1) + E\left[v\left(\frac{\rho S}{2} - \frac{x_1}{2} - \frac{1-\rho^2}{4}\right) + v\left(\frac{2W - \rho S}{2} + \frac{x_1}{2} + \frac{1-\rho^2}{4}\right)\right] - \lambda \log \sqrt{1-\rho^2}.$$

Since v is exponential, this objective function is supermodular in x_1 and ρ^2: savings in period 1 and signal precision in period 2 are substitutes. On one hand, higher savings in period 1 decrease the demand for information in period 2 due to an income effect. On the other hand, a lower ρ^2 in period 2 generates, in period 1, more uncertainty about the future, which increases savings by a prudent agent.

Since the optimal ρ^2 is decreasing in λ, the relation between x_1 and ρ^2 implies that attention costs generate a motive for precautionary savings also in period 1. Specifically, the agent's optimal consumption in period 1 is given by

$$x_1(\lambda) = -\frac{2-\rho^2(\lambda)}{12},$$

which is decreasing in λ. Moreover, the relation between x_1 and λ can be exploited to infer the agent's attention costs. From the optimality conditions for the consumption-saving problem:

$$\begin{aligned} x_1 = -\frac{1}{6} \quad &\Leftrightarrow \quad \lambda \geq \frac{1}{2}e^{\frac{1}{6}}, \\ x_1 \in \left(-\frac{1}{6}, -\frac{1}{12}\right) \quad &\Leftrightarrow \quad \lambda = -\frac{(12x_1+1)\,e^{-x_1}}{2}. \end{aligned}$$

As a result, when $x_1 > -\frac{1}{6}$, we can infer the unobservable attention cost parameter λ directly from the choice of x_1. When $x_1 = -\frac{1}{6}$, we can infer that the agent extracts an uninformative signal, and determine a lower bound on λ. Of course, these inferences are possible in the example only because attention costs are defined in terms of a single parameter (λ), and because we have made simplifying assumptions to solve the problem analytically. In the following section, we show how attention costs can be elicited in our framework for a general model of rational inattention.

4 Analysis

In this section we analyze rationally inattentive preferences. The results can be summarized as follows: Theorem 1 provides an axiomatic characterization, Proposition 1 shows when parameters can be uniquely identified, Theorem 2 provides an explicit procedure to elicit attention costs, Propositions 2-4 address comparative statics, and finally Propositions 5-7 characterize special cases.

4.1 Characterization

Let $\succsim$ be a rationally inattentive preference over menus. The following structural axioms are straightforward implications of Definition 1.

Axiom 1. *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. *For all menus F, G and H, the following sets are closed:*

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

Axiom 3. *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 $\alpha z + (1-\alpha)y \succ x$.*

Axiom 4. *For all menus F and G, acts h and h', and $\alpha \in (0,1)$,*

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

Axiom 1 states that $\succsim$ is complete and transitive, Axiom 2 is a mixture-continuity condition (Herstein and Milnor [1953]), and Axiom 3 states that is is always possible to find an arbitrarily good (or bad) outcome (see, e.g., Maccheroni, Marinacci, and Rustichini [2006, Lemma 29]). These axioms follow from Definition 1, since $\succsim$ has a real-valued, mixture-continuous, unbounded representation.

Axiom 4 states that, when the DM compares menus $\alpha F + (1-\alpha)h$ and $\alpha G + (1-\alpha)h$, her preference does not depend on which singleton menu will be realized with probability $(1-\alpha)$: since information is redundant for singletons, the optimal allocation of attention depends only on α, F and G, but does not change if h is exchanged for the alternative act h'. The axiom follows from Definition 1, since the DM is Bayesian and the attention costs are additively separable in the representation of $\succsim$.

The following axioms reflect the main behavioral implications of rational inattention in our framework. In particular, the axioms show how two key features of the rational inattention model translate into observable behavior: (i) the DM chooses the optimal act conditional on the information conveyed by signal realizations, and (ii) the DM chooses an optimal signal for each menu, to balance the benefit and costs of attention.

Desire for flexibility

A rationally inattentive agent would prefer a larger opportunity set in order to better adapt the choice of an alternative to the realization of her signal, and the following axiom formalizes this desire for flexibility.

Axiom 5. *For all menus F and acts g, $F \cup \{g\} \succsim F$, with $F \cup \{g\} \sim F$ if there is $f \in F$ such that $f(\omega) \succsim g(\omega)$ for all ω.*

The first part of this axiom ($F \cup \{g\} \succsim F$) states that the DM values the gain in flexibility from adding act g to menu F. For example, while there could be an act $f \in F$ that the DM would prefer to g based on her information today, she could acquire new information tomorrow which would lead her to choose g instead. Hence, it is possible that $F \cup \{g\} \succ F \succ \{g\}$. However, if there is already an act $f \in F$ that the DM prefers to g in *all* states, then no information the DM could obtain would lead her to choose g in the future. The second part of the axiom therefore excludes that $F \cup \{g\} \succ F$ under these circumstances.

The desire for flexibility reflects that the DM chooses the optimal act from her menu based on the signal realization she observes. Attention limitations under rational inattention are therefore related only to information about the state of fundamentals, and the DM can always ignore inferior alternatives in her opportunity set. This feature distinguishes rational inattention from models of decision-making in which the DM is "inattentive" to alternatives in her opportunity set. In such models, adding alternatives to an opportunity set can actually make the DM worse off, since additional alternatives can "distract" her attention from the optimal choice (see, e.g., Masatlioglu, Nakajima, and Ozbay [2012], Manzini and Mariotti [2014], Ortoleva [2013]).

Desire for early resolution of uncertainty

A rationally inattentive agent would prefer an early resolution of mixture uncertainty (facing menu F for sure) rather than a late resolution (facing menu $\alpha F + (1-\alpha)G$, when $F \sim G$), because she can then choose a signal optimal for the opportunity set that determines her final payoffs. The following axiom formalizes this desire for early resolution of uncertainty.

Axiom 6. *For all menus F and G, if $F \sim G$ then $F \succsim \alpha F + (1-\alpha)G$ for all $\alpha \in (0,1)$.*

In the mixed menu $\alpha F + (1-\alpha)G$ the DM chooses acts $f \in F$ and $g \in G$, not knowing whether the choice from F or G will actually determine her outcome. Since the optimal way to allocate attention may differ for menus F and G, the DM would like this additional uncertainty to be resolved before processing information about the state ω. As a result, if F is indifferent to G, the DM prefers F to the mixed menu $\alpha F + (1-\alpha)G$.[12]

The desire for early resolution of uncertainty reflects that the DM chooses an optimal signal for each menu, to balance benefits and costs of attention. In particular, $F \succ \alpha F + (1-\alpha)G$ (when $F \sim G$) implies that the optimal signal is not the same for menus F and G. The possibility of a strict desire for early resolution therefore distinguishes rational inattention from models of information acquisition in which the DM cannot adapt her information to the specific menu. In Section 4.5 we show how different restrictions on the DM's ability to allocate attention are revealed by her attitudes towards temporal resolution of uncertainty.

Representation theorem

The following result establishes that the preceding axioms identify all implications of rational inattention, providing a way to test if the DM is rationally inattentive in our framework:

Theorem 1. *A binary relation $\succsim$ over menus is a rationally inattentive preference if and only if it satisfies Axioms 1-6.*

4.2 Identification

We now consider identification of model parameters $(u, \bar{p}, c)$ in our framework. First we observe that the prior and utility function of the DM can be uniquely obtained from preferences over singleton menus, up to the usual positive affine transformations. We write $(u, \bar{p}) \approx (u', \bar{p}')$ if there are $\alpha > 0$ and $\beta \in \mathbb{R}$ such that $u = \alpha u' + \beta$ and $\bar{p} = \bar{p}'$.

Remark 1. Suppose $\succsim$ is a rationally inattentive preference represented by $(u, \bar{p}, c)$. Then the restriction of $\succsim$ over acts is represented by $(u, \bar{p})$, i.e., for all acts f and g,

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

Moreover, if $(u', \bar{p}')$ also represents the preference relation $\succsim$ over acts, then $(u, \bar{p}) \approx (u', \bar{p}')$.

To identify also the attention costs, we introduce some additional regularity conditions.

Paying no attention is costless

Let $\pi_0 \in \Pi(\bar{p})$ denote the Dirac distribution concentrated on the prior $\bar{p}$ (i.e., $\pi_0(\bar{p}) = 1$).

[12] To provide additional intuition for Axiom 6, we discuss the early resolution of uncertainty in an extension of our framework, and in the context of the consumption -saving example, in Appendix A.4.

Condition 1. $c(\pi_0) = 0$.

The signal π_0 conveys no new information relative to the prior. Condition 1 therefore states that if the DM does not pay attention to any information, she incurs no cost.

Blackwell monotonicity

To state the second condition, we recall the *Blackwell order*, which ranks signals by their "informativeness" (Blackwell [1953]):

Definition 2. Signal π is *more informative* than signal ρ, denoted $\pi \trianglerighteq \rho$, if

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

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

The Blackwell order has a natural interpretation in our context.

Remark 2. For a rationally inattentive DM, $\pi \trianglerighteq \rho$ if and only if, for all menus F, $b_F(\pi) \geq b_F(\rho)$, i.e., signal π is more informative than signal ρ if and only if the decision-making benefits of π are always higher than for ρ.

Condition 2. $\pi \trianglerighteq \rho$ implies $c(\pi) \geq c(\rho)$.

Condition 2 states that extracting a more informative signal requires a greater input of attention, and is therefore costlier.

Convexity

To state the last condition, we observe that $\Pi(\bar{p})$ is a convex set.

Condition 3. For signals π and ρ, and $\alpha \in (0, 1)$, $c(\alpha\pi + (1 - \alpha)\rho) \leq \alpha c(\pi) + (1 - \alpha)c(\rho)$.

The signal $\alpha\pi + (1 - \alpha)\rho$ can be interpreted as a randomization over the signals π and ρ (with weights α and $(1 - \alpha)$, respectively). Condition 3 states that the cost of the "mixed" signal $\alpha\pi + (1 - \alpha)\rho$ is lower than the expected cost $\alpha c(\pi) + (1 - \alpha)c(\rho)$.

Identifying attention costs

Conditions 1-3 are natural properties satisfied by the attention cost functions considered in the literature (for example, costs defined in terms of the Shannon information). We therefore regard these conditions as canonical:

Definition 3. An attention cost function c is *canonical* if it satisfies Conditions 1-3.

We can now state the identification result.

Proposition 1. *Suppose $\succsim$ is a rationally inattentive preference. Then $\succsim$ can be represented by some $(u, \bar{p}, c)$ where c is canonical. Moreover, if $(u', \bar{p}', c')$ also represent $\succsim$, and c' is canonical, then there exist some $\alpha > 0$ and $\beta \in \mathbb{R}$ such that $u = \alpha u' + \beta$, $\bar{p} = \bar{p}'$, and $c = \alpha c'$.*

Proposition 1 establishes that, if we restrict the rational inattention model to canonical attention costs, menu choice data allows a unique identification of parameters up to a standard normalization of utility. In particular, Proposition 1 identifies the properties of cost functions that can be assumed without loss of generality, having no additional behavioral implications, and demonstrates that any additional properties (such as the linearity property of costs based on the Shannon information in Example 1) will impose further restrictions on the behavior of the agent.

To illustrate why the identification may not be possible without these properties, consider signals π and ρ where π is more informative than ρ, but $c(\pi) < c(\rho)$ (i.e., c violates Condition 2). Since $\pi \trianglerighteq \rho$ the benefit of information from signal π is greater than from signal ρ for any menu. As a result, if we define a cost function c' that agrees everywhere with c, except that $c'(\rho) \in (c(\pi), c(\rho))$, then the attention cost function c' is observationally equivalent to c (because the signal ρ is never optimal under both cost functions). The restriction to canonical attention costs allows a unique identification by "removing" such redundancies.

4.3 Elicitation

We are now ready to formalize the intuition from the introduction, that a DM's "willingness-to-pay" for different menus can be used to construct her attention costs. To illustrate the construction, consider a rationally inattentive DM with preference relation $\succsim$ represented by V, with a given utility u and prior $\bar{p}$.[13] Now suppose we want to elicit the attention cost the DM incurs to extract a particular signal $\pi \in \Pi(\bar{p})$. Since signals are optimally chosen, for any menu F, $V(F) \geq b_F(\pi) - c(\pi)$. Using the certainty equivalent, this inequality can be rearranged as $c(\pi) \geq b_F(\pi) - u(x_F)$, which gives a lower bound on $c(\pi)$.[14] By varying the menu F, the lower bound can be increased to improve the approximation of $c(\pi)$. The following theorem establishes that, with sufficient data, this procedure gives the unique canonical attention cost for signal π.

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

$$c(\pi) = \sup_{F \in \mathbb{F}} \left[\int_{\Delta(\Omega)} \max_{f \in F} \left(\int_{\Omega} u(f(\omega))\, p(d\omega) \right) \pi(dp) - u\left(x_F\right) \right] \quad \forall \pi \in \Pi(\bar{p}),$$

[13]By Remark 1, the restriction of $\succsim$ on singletons menus identifies the utility and prior, and standard methods from expected utility theory can therefore be used to elicit these parameters.

[14]Axioms 1 and 2 imply that every menu F has a certainty equivalent $x_F \in X$, such that $F \sim x_F$.

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

Theorem 2 establishes that data on certainty equivalents for menus can be used to elicit the attention costs. Such data could be generated by enriching experiments already proposed in the literature to study implications of attention limitations. Consider, for example, the experimental design proposed by Caplin and Dean [2013, Section 6].[15] In their basic treatment, there are two equally likely states determined by the composition of 100 colored balls displayed on a screen: a "blue" state with 52 blue and 48 red balls, and a "red" state with 52 red and 48 blue balls. Before the screen is revealed, participants are given a menu of acts (each paying monetary outcomes conditional on the state) and are informed that they are to choose an act after the screen is displayed to them. Participants therefore face an attention allocation problem: by inspecting the screen they can extract information about the state to choose a better act from their menu.[16] To implement our elicitation procedure, one can complement this experimental setting by first asking the participants for the sure amount of money they are willing to trade for different menus *before* the screen is displayed. Standard methods can be used to obtain such willingness-to-pay data in an incentive compatible way (e.g., Becker, DeGroot, and Marschak [1964]).

4.4 Comparative Statics

We now study how structural changes in the model correspond to changes in behavior. We compare two rationally inattentive decision-makers, DM1 and DM2, with preference relations $\succsim_1$ 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.

Attention costs

Since attention costs are the novel parameter in rational inattention models, we first ask how a change in attention costs translates into observable changes in behavior.

Definition 4. DM1 has *lower attention costs* than DM2 if $(u_1, \bar{p}_1) = (u_2, \bar{p}_2)$ and $c_1 \leq c_2$.[17]

Our first comparative static result indicates that a decrease in attention costs corresponds to a greater desire for flexibility.

Proposition 2. *The following conditions are equivalent.*

[15]Gabaix et al. [2006] and Cheremukhin et al. [2014] also propose experiments on attention, which could be adapted to generate the willingness-to-pay data used in our elicitation procedure.

[16]Caplin and Dean [2013] present participants with each menu 50 times, and use the data to approximate a state-dependent stochastic choice, to study how participants respond to the attention allocation problem. Even though there are no time limits to inspect the screen, they find substantial evidence of inattention: subjects choose an act that delivers a lower monetary payoff in 35% of trials.

[17]The restriction to a common prior and utility is necessary to compare attention costs because the prior defines the domain of these costs, and the utility function defines the units in which they are measured.

(i) DM1 has lower attention costs than DM2.

(ii) For all menus F and acts g, $g \succ_1 F$ implies $g \succ_2 F$.

The preference $g \succ F$ 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 g. Condition (ii) therefore reflects that DM1 has a stronger desire for flexibility than DM2. Proposition 2 shows that this change in behavior reveals that DM1 has lower attention costs than DM2, since with lower attention costs DM1 is better able to exploit the flexibility in a menu.

Reallocation of attention

A second key feature of the rational inattention model is that the DM reallocates attention depending on what information is most relevant for her (see, e.g., Kacperczyk et al. [2012, p.2]). Formally, we denote by $\partial V(F)$ the set of optimal signals given menu F,

$$\partial V(F) := \arg \max_{\pi \in \Pi(\bar{p})} \left[b_F(\pi) - c(\pi) \right],$$

and say that the DM reallocates attention between two menus F and G when $\partial V(F) \cap \partial V(G) = \emptyset$.

Definition 5. DM1 has a *greater tendency to reallocate attention* than DM2 if, for all menus F and G, $\partial V_2(F) \cap \partial V_2(G) = \emptyset$ implies $\partial V_1(F) \cap \partial V_1(G) = \emptyset$.

Our next comparative static result indicates that an increased tendency to reallocate attention corresponds to a greater desire for early resolution.

Proposition 3. *The following conditions are equivalent.*

(i) DM1 has a greater tendency to reallocate attention than DM2.

(ii) For all menus F and G with certainty equivalents x_F^i and x_G^i ($i = 1, 2$), and $\alpha \in (0, 1)$,

$$\alpha x_F^2 + (1-\alpha) x_G^2 \succ_2 \alpha F + (1-\alpha) G \quad \Rightarrow \quad \alpha x_F^1 + (1-\alpha) x_G^1 \succ_1 \alpha F + (1-\alpha) G.$$

For a rationally inattentive preference $\alpha x_F + (1-\alpha) x_G$ is always weakly preferred to $\alpha F + (1-\alpha) G$. Indeed, in the presence of Axioms 1-5, the condition $\alpha x_F + (1-\alpha) x_G \succsim \alpha F + (1-\alpha) G$ for all menus F and G, and $\alpha \in (0, 1)$, is equivalent to Axiom 6. Condition (ii) therefore reflects that DM1 has a stronger desire for early resolution than DM2. Proposition 3 shows that this change in behavior reveals that DM1 has a greater tendency to reallocate attention than DM2, since with a greater tendency to reallocate attention DM1 is more often affected by late resolution of uncertainty.

Attentiveness

Finally, we characterize the behavioral implications of a comparison of "attentiveness":

Definition 6. DM1 is *more attentive* than DM2 when, given $(u_1, \bar{p}_1) \approx (u_2, \bar{p}_2)$, for every menu F and signal $\pi_2 \in \partial V_2(F)$, there is a signal $\pi_1 \in \partial V_1(F)$ such that $\pi_1 \trianglerighteq \pi_2$.

Intuitively, extracting a signal which is more informative in the sense of Blackwell [1953] requires more attention. Accordingly by Definition 6, DM1 is said to be more attentive than DM2 if, for every menu, she extracts a more informative signal. Our last comparative static result identifies this increase in attentiveness with a trade-off between desire for flexibility and early resolution of uncertainty.

Proposition 4. *The following conditions are equivalent.*

(i) DM1 is more attentive than DM2.

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

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

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 late resolution in $\alpha F + (1-\alpha)G$ does not allow the DM to allocate attention specific to the menu that determines her final payoffs. As such, the DM faces a trade-off between flexibility and temporal resolution. The pattern $\alpha F + (1-\alpha)h \succ \alpha F + (1-\alpha)G$ indicates that the DM would prefer the early resolution to the gain in flexibility. Condition (ii) therefore reflects that DM1 is both better able to exploit flexibility, and less affected by late resolution of uncertainty, than DM2. Proposition 4 shows that this change in behavior reveals that DM1 is more attentive than DM2, since with more informative signals DM1 is both better able to exploit flexibility and less affected by late resolution of uncertainty.

4.5 Special cases

Finally, to illustrate how different restrictions on attention allocation are revealed by attitudes toward temporal resolution of uncertainty, we consider some special cases of the general rational inattention model: (i) a model in which attention is constrained rather than costly; (ii) a model in which the DM can acquire a specific signal, at a cost, or remain uninformed; and (iii) a model in which the signal is exogenously given and does not vary with the opportunity set.

Constrained attention

In a model of constrained attention, the DM has a limited amount of attention to allocate, but otherwise incurs no cost of attention. The key parameter to formalize this type of attention limitation is an *attention constraint set*, which is a non-empty closed set of signals $\Gamma \subset \Pi(\bar{p})$. Given such a set Γ, the DM chooses signal $\pi \in \Gamma$ to maximize the benefit of information for her menu.

Definition 7. A binary relation $\succsim$ over menus is a *constrained attention preference* if it is represented by a functional $V : \mathbb{F} \to \mathbb{R}$, defined by

$$V(F) = \max_{\pi \in \Gamma} \left(\int_{\Delta(\Omega)} \max_{f \in F} \left(\int_{\Omega} u(f(\omega))\, p(d\omega) \right) \pi(dp) \right) \qquad \forall F \in \mathbb{F},$$

where $u : X \to \mathbb{R}$ is an unbounded affine utility function, $\bar{p} \in \Delta(\Omega)$ is a prior, and $\Gamma \subset \Pi(\bar{p})$ is an attention constraint set.

Example 2. Sims [2003] initially formulated rational inattention in terms of a constraint set $\Gamma = \{\pi : I(\pi) \leq \kappa\}$, where $\kappa \geq 0$. In the interpretation, the preference parameter κ represents a capacity constraint on information gain, measured by the Shannon information I.

Constrained attention is a special case of rational inattention, in which the cost function only takes values in $\{0, \infty\}$. Following Sims [2003], such constrained models have been widely used in applications (see, e.g., Wiederholt [2010]). However, as argued by Sims [2010, p. 29], it is "more appealing to think of people as applying a small part of their full information processing capacity to monitoring economic signals, with a stable shadow price on that processing capacity, than to suppose that they have a fixed capacity constraint". In our framework, such attention constraints are revealed by attitudes towards temporal resolution of uncertainty, providing a way to evaluate how capacity constraints restrict the behavior of rationally inattentive agents.

Axiom 7. *For all menus F and acts h, if $F \sim h$ then $F \sim \alpha F + (1-\alpha)h$ for all $\alpha \in (0,1)$.*

In the constrained attention model, the optimal allocation of attention for menu F is also optimal for menu h (since h is a singleton menu and there is no cost of attention). As a result, the DM's allocation of attention does not depend on whether menu F or h will determine her final payoff, and she therefore does not have a strict desire for early resolution of uncertainty for such mixtures. In particular, the DM's "input" of attention for $\alpha F + (1-\alpha)h$ does not change as α varies, even though for an α close to 1 information is more valuable than for an α close to 0. The following proposition establishes that Axiom 7 identifies the additional restriction of constrained attention in our framework.

Proposition 5. *Let $\succsim$ be a rationally inattentive preference. Then $\succsim$ is a constrained attention preference if and only if it satisfies Axiom 7.*

Binary attention

We now consider a model in which the DM faces a binary information acquisition problem: she can decide whether to acquire an informative signal π_1 at a cost, or remain uninformed (acquiring the uninformative signal π_0):

Definition 8. A binary relation $\succsim$ over menus is an *binary attention preference* if it is represented by a functional $V : \mathbb{F} \to \mathbb{R}$, defined by

$$V(F) = \max_{\pi \in \{\pi_0, \pi_1\}} \left[\int_{\Delta(\Omega)} \max_{f \in F} \left(\int_{\Omega} u(f(\omega))\, p(d\omega) \right) \pi(dp) - c(\pi) \right] \qquad \forall\, F \in \mathbb{F}$$

where $u : X \to \mathbb{R}$ is an unbounded affine utility function , $\bar{p} \in \Delta(\Omega)$ is a prior, $\pi_1 \in \Pi(\bar{p})$ is a signal, and c is a canonical attention cost function.

Binary attention is another special case of rational inattention, in which the DM is restricted to choose between an informative signal π_1 or the uninformative signal π_0. Such models are common in the literature on information acquisition, except that the cost of information is not interpreted in terms of attention. For example, in the seminal paper of Grossman and Stiglitz [1980], investors are uncertain about the return of an asset, and can acquire at a monetary cost, a specific random variable conveying information about the asset returns. The main restriction in these models is that there is only one informative signal and, in our framework, such restrictions on the set of available signals are revealed by the DM's attitude towards temporal resolution of uncertainty:

Axiom 8. *For all menus F and G, if $F \sim G$, $F \succ f$ for all $f \in F$ and $G \succ g$ for all $g \in G$ then $F \sim \alpha F + (1-\alpha)G$ for all $\alpha \in (0,1)$.*

The condition $F \succ f$ for all $f \in F$ and $G \succ g$ for all $g \in G$ indicates that the DM acquires an informative signal for each menu (since she strictly values the flexibility offered by each menu). However, since there is only one informative signal available, the optimal allocation of attention must be the same for menus F and G. As a result, how the DM allocates attention does not depend on whether menu F or G will determine the final payoff, and she therefore does not have a strict desire for early resolution of uncertainty for such mixtures. The following proposition establishes that Axiom 8 identifies the additional restriction of binary attention in our framework.

Proposition 6. *Let $\succsim$ be a rationally inattentive preference. Then $\succsim$ is an binary attention preference if and only if it satisfies Axiom 8.*

Exogenous attention

Finally, we consider a model in which the DM receives an exogenous signal π_1 at no cost, and cannot vary the allocation of attention with her menu:

Definition 9. A binary relation $\succsim$ over menus is an *exogenous attention preference* if it is represented by a functional $V : \mathbb{F} \to \mathbb{R}$, defined by

$$V(F) = \int_{\Delta(\Omega)} \max_{f \in F} \left(\int_{\Omega} u(f(\omega))\, p(d\omega) \right) \pi_1(dp) \qquad \forall\, F \in \mathbb{F},$$

where $u : X \to \mathbb{R}$ is an unbounded affine utility function , $\bar{p} \in \Delta(\Omega)$ is a prior, and $\pi_1 \in \Pi(\bar{p})$ is an exogenous signal.

Exogenous attention is a special case of both constrained and binary attention: the DM can choose only the signals π_0 and π_1 (binary attention), and both signals are costless (constrained attention). A recent literature has argued that the distinction between exogenous attention and endogenous attention (as in rational inattention, where the DM *chooses* an optimal signal) can have important implications for policy, but also that the behavioral implications of these models can be difficult to distinguish (see, e.g., Paciello and Wiederholt [2014] and Kacperczyk, Van Nieuwerburgh, and Veldkamp [2012]). This model was first characterized by Dillenberger, Lleras, Sadowski, and Takeoka [2013], who interpret it as a model of subjective learning. We provide an alternative characterization, to show how the distinction between endogenous and exogenous attention can be revealed by attitudes towards temporal resolution of uncertainty:

Axiom 9. *For all menus F and G, if $F \sim G$ then $F \sim \alpha F + (1-\alpha)G$ for all $\alpha \in (0, 1)$.*

In the exogenous attention model, the optimal allocation of attention does not vary. As a result, whether menu F or G will determine the DM's final payoff does not affect how the DM allocates her attention, and she therefore does not have any strict desire for early resolution of uncertainty. The following proposition establishes that Axiom 9 identifies the additional implication of exogenous attention in our framework.

Proposition 7. *Let $\succsim$ be a rationally inattentive preference. Then $\succsim$ is an exogenous attention preference if and only if it satisfies Axiom 9.*

5 Discussion

Our formal analysis of rational inattention is informed by several strands of the decision theory literature.

Flexibility

The decision-theoretic analysis of menu-choice was initiated by Kreps [1979] to formalize a desire for flexibility. In such models, the DM prefers larger opportunity sets because they allow her to better adapt the choice of an alternative to "subjective" contingencies which realize in the future.

While the existing literature (e.g., Dekel, Lipman, and Rustichini [2001]) has mostly focused on contingencies affecting the DM's tastes, the effect of such contingencies are excluded under rational inattention. For a rationally inattentive DM the relevant contingencies are realizations of signals, that convey information and affect her beliefs about the state ω. This particular interpretation for subjective contingencies is reflected in our Axiom 5, which implies that the DM is strategically rational (Kreps [1988, p. 88]) on menus consisting only of constant acts ($F \subset X$).

Temporal resolution of uncertainty

The decision-theoretic analysis of attitudes towards temporal resolution of uncertainty was initiated by Kreps and Porteus [1978]. In such models, a desire for early resolution can be driven by a "hidden" decision that the DM has to make before all uncertainty is resolved (Machina [1984]). In the rational inattention model, the particular hidden decision problem facing the DM is the extraction of an optimal signal, which induces a desire for early resolution of uncertainty in our framework (Axiom 6).

Variational preferences

For our formal analysis, we start with the observation that the choice of a signal can be seen as a variational problem. Given a menu F, a rationally inattentive DM chooses a signal $\pi \in \Pi(\bar{p})$ to maximize the value of information $b_F(\pi) - c(\pi)$, which is concave and upper-semicontinuous in π, and therefore corresponds to a variational problem. This observation provides a formal connection between our work and Maccheroni, Marinacci, and Rustichini [2006]. To study how different models of ambiguity are related, they introduce a class of preferences over acts whose representation also involves a variational problem. This common variational structure allows us to adopt some of the ideas and tools developed by Maccheroni, Marinacci, and Rustichini [2006] to our framework.

Ergin and Sarver [2010] also introduce this variational structure to a menu-choice framework, where menus are sets of lotteries (in particular there is no state of fundamentals). In their framework, they characterize a model of costly contemplation, in which the DM is uncertain about her future tastes (e.g., risk preferences) and is able to exert costly contemplation to better understand them before choosing a lottery. As discussed above, in our axiomatic characterization, uncertainty about tastes is excluded by Axiom 5. As a result, we are able to focus on "costly contemplation" about the state ω, and study the menu-choice implications of rational inattention.

Stochastic choice

A number of papers (e.g., Caplin and Dean [2013], Ellis [2012], Matejka and McKay [2011]) have recently followed an alternative approach to study rational inattention from a decision-theoretic perspective. These papers look at the alternatives chosen by a rationally inattentive DM from different menus after processing information about the state. Such choices, being contingent on the signal realizations, stochastically depend on the realized state. This alternative approach can be viewed as complementary to ours. For instance, as discussed in Section 4.3, in experimental settings both types of data could be jointly collected and complement each other. However, peculiar to our approach is the precise identification of attention costs for a general model of rational inattention.

6 Conclusion

Rational inattention has become an influential approach to study implications of attention limitations in economic environments. The key parameter in these models is the cost of attention, which is essential for predicting the behavior of the agent. However, these costs are difficult to measure as they can incorporate many factors that are not directly observable (such as time, effort and cognitive resources).

In this paper, we propose a framework to identify attention costs from choice data, by characterizing the implications of rational inattention for choices over menus. Our approach takes the perspective of the DM when she chooses a menu today, with the prospect of processing information tomorrow before selecting an alternative. The preferences over menus reveal the trade-off the DM faces when she balances the benefits of information (which depend on the menu) against the cost of attention (which is a characteristic of the individual). We provide an explicit procedure to elicit attention costs in this framework, and show how the choice-data can be obtained in dynamic environments (e.g., a consumption-saving problem) or generated in experimental settings.

Our axiomatic characterization provides a way to test if the DM is rationally inattentive, and shows that key features of rational inattention models can be revealed by the DM's attitudes towards flexibility and temporal resolution of uncertainty. For example, we show that special cases of the model - where attention is constrained, binary or exogenous - can be compared by observing the DM's attitude towards temporal resolution of uncertainty. Moreover, comparisons on attention costs, tendency to reallocate attention, and attentiveness, are related to intuitive changes in attitudes towards flexibility and temporal resolution of uncertainty.

As the growing literature attests, rational inattention has the potential to explain a wide-range of economic phenomena, and empirical analysis of these models is therefore an important area for future research. By providing a way to elicit attention costs, perform comparative statics, and evaluate different models of attention limitations with choice data, our analysis can support research in this direction.

References

Aliprantis, C. D. and K. C. Border (2006): *Infinite Dimensional Analysis*, Springer, third ed.

Anscombe, F. J. and R. J. Aumann (1963): "A Definition of Subjective Probability," *Annals of Mathematical Statistics*, 34, 199–205.

Becker, G. M., M. H. DeGroot, and J. Marschak (1964): "Measuring utility by a single-response sequential method," *Behavioral Science*, 9, 226–232.

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

Brown, J., T. Hossain, and J. Morgan (2010): "Shrouded Attributes and Information Suppression: Evidence from the Field," *The Quarterly Journal of Economics*, 125, 859–876.

Caplin, A. and M. Dean (2013): "Rational Inattention and State Dependent Stochastic Choice," .

Cerreia-Vioglio, S., F. Maccheroni, M. Marinacci, and A. Rustichini (2012): "Niveloids and Their Extensions: Risk Measures on Small Domains," Working Paper 458, IGIER, Bocconi University.

Cheremukhin, A., A. Popova, and A. Tutino (2014): "A Theory of Discrete Choice with Information Costs," Mimeo.

Chetty, R., A. Looney, and K. Kroft (2009): "Salience and Taxation: Theory and Evidence," *American Economic Review*, 99, 1145–77.

Conlisk, J. (1996): "Why Bounded Rationality?" *Journal of Economic Literature*, 34, 669–700.

Dekel, E., B. L. Lipman, and A. Rustichini (2001): "Representing Preferences with a Unique Subjective State Space," *Econometrica*, 69, 891–934.

DellaVigna, S. (2009): "Psychology and Economics: Evidence from the Field," *Journal of Economic Literature*, 47, 315–72.

DellaVigna, S. and J. M. Pollet (2009): "Investor Inattention and Friday Earnings Announcements," *Journal of Finance*, 64, 709–749.

Dillenberger, D., J. S. Lleras, P. Sadowski, and N. Takeoka (2013): "A Theory of Subjective Learning," .

Dreze, J. H. and F. Modigliani (1972): "Consumption decisions under uncertainty," *Journal of Economic Theory*, 5, 308–335.

Ellis, A. (2012): "Foundations for Optimal Inattention," .

Ergin, H. and T. Sarver (2010): "A Unique Costly Contemplation Representation," *Econometrica*, 78, 1285–1339.

Gabaix, X. and D. Laibson (2006): "Shrouded Attributes, Consumer Myopia, and Information Suppression in Competitive Markets," *The Quarterly Journal of Economics*, 121, 505–540.

Gabaix, X., D. Laibson, G. Moloche, and S. Weinberg (2006): "Costly Information Acquisition: Experimental Analysis of a Boundedly Rational Model," *American Economic Review*, 96, 1043–1068.

Gollier, C. (2004): *The Economics of Risk and Time*, MIT Press.

Grossman, J. and J. E. Stiglitz (1980): "On the impossibility of informationally efficient markets," *American Economic Review*, 393–408.

Herstein, I. N. and J. Milnor (1953): "An Axiomatic Approach to Measurable Utility," *Econometrica*, 21, 291–297.

Hirshleifer, D., S. S. Lim, and S. H. Teoh (2009): "Driven to Distraction: Extraneous Events and Underreaction to Earnings News," *The Journal of Finance*, 64, 2289–2325.

Kacperczyk, M., S. Van Nieuwerburgh, and L. Veldkamp (2012): "Rational Attention Allocation over the Business Cycle," .

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

Kimball, M. S. (1990): "Precautionary Saving in the Small and in the Large," *Econometrica*, 58, 53–73.

Kreps, D. M. (1979): "A Representation Theorem for "Preference for Flexibility"," *Econometrica*, 47, 565–577.

——— (1988): *Notes on the Theory of Choice*, Westview Press.

Kreps, D. M. and E. L. Porteus (1978): "Temporal Resolution of Uncertainty and Dynamic Choice Theory," *Econometrica*, 46, pp. 185–200.

Lipman, B. L. (1995): "Information Processing and Bounded Rationality: A Survey," *Canadian Journal of Economics*, 28, 42–67.

Luo, Y. (2010): "Rational Inattention, Long-run Consumption Risk, and Portfolio Choice," *Review of Economic Dynamics*, 13, 843–860.

Maccheroni, F., M. Marinacci, and A. Rustichini (2006): "Ambiguity Aversion, Robustness, and the Variational Representation of Preferences," *Econometrica*, 74, 1447–1498.

MACHINA, M. (1984): "Temporal risk and the nature of induced preferences," *Journal of Economic Theory*, 33, 199–231.

MAĆKOWIAK, B. AND M. WIEDERHOLT (2009): "Optimal Sticky Prices under Rational Inattention," *American Economic Review*, 99, 769–803.

——— (2011): "Business Cycle Dynamics under Rational Inattention," Working Paper 1331, ECB.

MANZINI, P. AND M. MARIOTTI (2014): "Imperfect Attention and Menu Evaluation," .

MASATLIOGLU, Y., D. NAKAJIMA, AND E. Y. OZBAY (2012): "Revealed Attention," *American Economic Review*, 102, 2183–2205.

MATĚJKA, F. (2012): "Rationally Inattentive Seller: Sales and Discrete Pricing," .

MATEJKA, F. AND A. MCKAY (2011): "Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model," CERGE-EI Working Papers wp442, The Center for Economic Research and Graduate Education - Economic Institute, Prague.

——— (2012): "Simple Market Equilibria with Rationally Inattentive Consumers," *American Economic Review*, 102, 24–29.

MILGROM, P. AND I. SEGAL (2002): "Envelope Theorems for Arbitrary Choice Sets," *Econometrica*, 70, 583–601.

ORTOLEVA, P. (2013): "The price of flexibility: Towards a theory of Thinking Aversion," *Journal of Economic Theory*, 148, 903–934.

PACIELLO, L. AND M. WIEDERHOLT (2014): "Exogenous Information, Endogenous Information, and Optimal Monetary Policy," *Review of Economic Studies*, 81, 356–388.

ROCKAFELLAR, R. T. (1970): *Convex Analysis*, Princeton University Press.

——— (1974): *Conjugate Duality and Optimization*, Society for Industrial and Applied Mathematics.

RUSZCZYŃSKI, A. AND A. SHAPIRO (2006): "Optimization of Convex Risk Functions," *Mathematics of Operations Research*, 31, 433–452.

SARVER, T. (2012): "Optimal Reference Points and Anticipation," .

SIMS, C. A. (1998): "Stickiness," *Carnegie-Rochester Conference Series on Public Policy*, 49, 317–356.

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

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

——— (2010): "Rational Inattention and Monetary Economics," Elsevier, vol. 3 of *Handbook of Monetary Economics*, 155 – 181.

Tutino, A. (2012): "Rationally Inattentive Consumption Choices," *Review of Economic Dynamics*, forthcoming.

van Nieuwerburgh, S. and L. Veldkamp (2010): "Information Acquisition and Under-Diversification," *Review of Economic Studies*, 77, 779–805.

Veldkamp, L. L. (2011): *Information Choice in Macroeconomics and Finance*, Princeton University Press.

Wiederholt, M. (2010): "Rational Inattention," in *The New Palgrave Dictionary of Economics*, ed. by S. N. Durlauf and L. E. Blume, Palgrave Macmillan.

Yang, M. (2013): "Optimality of Debt under Flexible Information Acquisition," .

A Appendix

In this section we prove the results (theorems and propositions) stated in Section 4 of the paper.

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 by $ca(\Delta(\Omega))$ the linear space of signed measures of bounded variation on $\Delta(\Omega)$ (Aliprantis and Border [2006, p. 399]). For each $\pi \in ca(\Delta(\Omega))$ 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 $ca(\Delta(\Omega))$ with the weak* topology. Therefore $ca(\Delta(\Omega))$ 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 $ca(\Delta(\Omega))$ (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 $\alpha \in \Psi$; *monotone* if $V(\varphi) \geq V(\psi)$ for all $\varphi, \psi \in \Psi$ such that $\varphi \geq \psi$; *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 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, Maccheroni, Marinacci, and Rustichini [2012], 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,

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

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,

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

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),

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

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 DM, $u(X)$ is unbounded. The following additional properties of $\Phi_{\mathcal{F}}$ and $\Phi_{\mathbb{F}}$ holds when $u(X)$ is unbounded above (analogous properties hold when $u(X)$ is unbounded below):

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

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

(iii) $0 \in u(X)$ implies $\alpha\varphi_F \subset \Phi_{\mathbb{F}}$ 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) $\Phi_{\mathbb{F}} + \mathbb{R}$ *is dense in* Φ.

A.2 Proofs of the Results in the Paper

We can now prove the results in Section 4. For simplicity, we assume that utility functions are unbounded above (the case where they are unbounded below is analogous and omitted).

Lemma 1. *Suppose that $\succsim$ satisfies Axioms 1-6. Then it is a rationally inattentive preference represented by some $(u, \bar{p}, c)$. Moreover, we can choose c to be canonical.*

Proof of Lemma 1

Assume that $\succsim$ satisfies Axioms 1-6. We start by showing an implication of Axiom 5 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 $F = \{f_1, \dots, f_n\}$ and $G = \{g_1, \dots, g_m\}$. By Axiom 5,

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

Again by Axiom 5,

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

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 the best outcome and y be the 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*

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

Proof. Take $f, g \in \mathcal{F}$ and fix $\alpha \in [0,1]$. Suppose that $f \sim g$: by Claim 2 we can choose $x \in X$ such that $x \sim f \sim g$. By Axiom 6 we have that $f \sim x \succsim \alpha g + (1-\alpha)f$ and $f \sim x \succsim \alpha f + (1-\alpha)g$. The first relation implies that $\alpha f + (1-\alpha)g \succsim g$ by Axiom 4. Since $f \sim g$, we conclude that $f \sim \alpha f + (1-\alpha)g$. The remainder of the proof then follows from Maccheroni, Marinacci, and Rustichini [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 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. 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 $\varphi_F = \varphi_G$ 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 $u \circ (\operatorname{co} F)$. By affinity of u, $\operatorname{co}(u \circ F) = u \circ (\operatorname{co} F)$, so that

$u \circ (\operatorname{co} F)$ is convex, closed and bounded. Let $E = \{e \in \mathbb{R}^\Omega : e \geq u \circ g\}$, then E is closed convex cone. Clearly, $u \circ (\operatorname{co} F)$ 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

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

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

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

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 $F \succsim G$.*

Proof. Let $G = \{g_1, \ldots, g_n\} \subset \operatorname{co} F$. For all $i = 1, \ldots, n$ we can write each $g_i = \sum_{j=1}^{m_i} \alpha_j^i f_j^i$ for $\alpha_1^i, \ldots, \alpha_{m_i}^i \geq 0$ summing up to one, and $f_1^i, \ldots, f_{m_i}^i \in F$. 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 Claim 1 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 $F \sim \sum_{k=1}^{l-1} \frac{\beta_k}{1-\beta_l} F$. Therefore by Axiom 6 $F \succsim \sum_{k=1}^l \beta_k F$. Since $F \subset \sum_{k=1}^l \beta_k F$, by Claim 1 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. Convexity of V follows easily from Axiom 6, Claims 2 and 3. 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 4, the obvious adaptation of the argument in Maccheroni, Marinacci, and Rustichini [2006, Proof of Lemma 28] provides that whenever α belongs to $u(X)$ we have for any $\varphi_F \in \Phi_{\mathbb{F}}$,

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

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

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

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

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

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

Since V is translation invariant on $\Phi_{\mathbb{F}}$, we can extend V uniquely to $\Phi_{\mathbb{F}} + \mathbb{R}$ by defining $V(\varphi) = V(\varphi + \alpha) - \alpha$ for any $\varphi \in \Phi_{\mathbb{F}} + \mathbb{R}$ and $\alpha \in \mathbb{R}$ such that $\varphi + \alpha \in \Phi_{\mathbb{F}}$. 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 $\Phi_{\mathbb{F}}$. □

Claim 7. For each menu F, $V(\varphi_F) = \max_{\pi \in \Pi(\bar{p})} \langle \varphi_F, \pi \rangle - c(\pi)$ where $c : \Pi(\bar{p}) \to \mathbb{R} \cup \{\infty\}$ is such that $c(\pi) = \sup_{F \in \mathbb{F}} \langle \varphi_F, \pi \rangle - V(\varphi_F)$ for all $\pi \in \Pi(\bar{p})$.

Proof. Since $\Phi_{\mathbb{F}}$ is convex and V is a convex niveloid, there is a real-valued functional W defined on $C(\Delta(\Omega))$ which 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 $\pi \in ca(\Delta(\Omega))$ such that

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

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

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

Thus, for all φ_F and π, that $V^*(\pi) \geqslant \langle \varphi_F, \pi \rangle - V(\varphi_F)$ and hence $V(\varphi_F) \geqslant \langle \varphi_F, \pi \rangle - V^*(\pi)$. Moreover, 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,

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

Now 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 $\Pi(\bar{p})$. □

We complete the proof by showing that c is a canonical attention cost function.

Claim 8. *The function c is a canonical attention cost function.*

Proof. Since V is normalized and $0 \in \Phi_{\mathbb{F}}$, $\langle 0, \pi \rangle - V(0) = 0$, it follows that $c \geq 0$. Moreover, since c is the pointwise supremum of a family of continuous affine functions, c is lower semi-continuous and convex, and since $\Phi_{\mathbb{F}} \subset \Phi$ it follows that $c(\pi) \geq c(\rho)$ whenever $\pi \trianglerighteq \rho$. Finally, observe that

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

By Axiom 5, F is preferred to any $f \in F$. Therefore $c(\pi_0) \leq 0$, and hence $c(\pi_0) = 0$. □

Proof of Theorem 1

It is straightforward to prove that a rationally inattentive preference $\succsim$ satisfies Axioms 1-6, and the intuition for this result is given in Section 4.1. The proof of the converse implication follows directly from Lemma 1.

Proof of Proposition 1

The proof that there always exists a canonical attention cost function follows directly from Lemma 1. Now 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 (Remark 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 $\pi \in \Pi(\bar{p})$,

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

Proof of Theorem 2

Let $(u, \bar{p}, c)$ represent a rationally inattentive preference, and let c be canonical. Define the functional $V : \Phi \to \mathbb{R}$ such that

$$V(\varphi) = \max_{\pi \in \Pi(\bar{p})} \langle \varphi, \pi \rangle - c(\pi) \quad \forall \varphi \in \Phi. \tag{1}$$

Fix $\pi \in \Pi(\bar{p})$. From Equation 1 we see that $c(\pi) \geq \langle \varphi_F, \pi \rangle - V(\varphi_F)$ for each menu F, and so $c(\pi) \geq \sup_{F \in \mathbb{F}} \langle \varphi_F, \pi \rangle - V(\varphi_F)$. We 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 $c(\pi) > \alpha \geq 0$. In order to do so, fix $0 \leq \alpha < c(\pi)$, and define the sets

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

where $\operatorname{epi} c$ is the epigraph of c. Since c is convex, 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 [2012, 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 $(0, 0) \in ca(\Delta(\Omega)) \times \mathbb{R}$ *is not a limit point of* $\operatorname{epi} c - \Phi^*_{\pi,\alpha}$.

Proof. Define the sets

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

Since $c \geq 0$, we have that $\operatorname{epi} c = A \cup B$, so that $\operatorname{epi} c - \Phi^*_{\pi,\alpha}$ is covered by $A - \Phi^*_{\pi,\alpha}$ and $B - \Phi^*_{\pi,\alpha}$. It is clear that $B - \Phi^*_{\pi,\alpha}$ is a subset of $ca(\Delta(\Omega)) \times [1, \infty)$, 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 $\operatorname{epi} c$ 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 $\operatorname{epi} c - \Phi^*_{\pi,\alpha}$ and $\operatorname{epi} c - \Phi^*_{\pi,\alpha}$ 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

$$\langle \varphi, \rho_1 \rangle - \langle \varphi, \pi + \rho_2 \rangle + \gamma(\beta - \alpha) \leq \kappa < \langle \varphi, 0 \rangle + \gamma(0) = 0 \tag{2}$$

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

Claim 10. *In expression (2), $\gamma \leq 0$ and φ is convex.*

Proof. Suppose, for contradiction, that $\gamma > 0$. Then we can take $\rho_2 = 0$, fix some $\rho_1 \in \operatorname{dom} c$, let β go to infinity and contradict (2).

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 $(\rho_1, \beta) \in \operatorname{epi} c$, 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 (2). Since $\langle \varphi, \rho_2 \rangle \geq 0$ for each $\rho_2 \in \Phi^*$ and $\Phi = \Phi^{**}$, we conclude that φ is convex (see Section A.1). □

We conclude the proof by showing that expression (2) 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 (2) . Then $\sup_{\varphi \in \Phi} \langle \varphi, \pi \rangle - V(\varphi) \geq \alpha$.*

Proof. Define $\psi = -\frac{\varphi}{\gamma}$. By Claim 10 φ is convex and $-\gamma > 0$ by hypothesis: hence $\psi \in \Phi$. From expression (2),

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

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 (2). Then* $\sup_{\psi\in\Phi}\langle\psi,\pi\rangle - V(\psi) \geq \alpha$.

Proof. Substituting $\rho_2 = 0$ in (2), 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

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

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 Proposition 2

Without loss of generality, we assume that $(u_1,\bar{p}_1) = (u_2,\bar{p}_2)$. 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 menu F, choose g so that $V_2(F) = V_2(g)$. Then it has to be the case that $V_1(F) \geq V_1(g) = V_2(F)$: hence $V_1 \geq V_2$.

On the other hand, suppose that $V_1 \geq V_2$. Choose a menu F and an act g: whenever $V_2(F) \geq V_2(g)$, 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. *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 up to one, the following statements are equivalent:*

(i) $\alpha_1 V(F_1) + \ldots + \alpha_n V(F_n) = V(\alpha_1 F_1 + \ldots + \alpha_n F_n)$;

(ii) $\partial V(\alpha_1 F_1 + \ldots + \alpha_n F_n) \subset \partial V(F_i)$ *for all* $i = 1,\ldots,n$;

(iii) $\partial V(F_1) \cap \ldots \cap \partial V(F_n) \neq \varnothing$.

Proof of Lemma 2

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

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

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

$$\alpha_1 V(F_1) + \ldots + \alpha_n V(F_n) = V\left(\alpha_1 F_1 + (1-\alpha_1)G\right) \leq \alpha_1 V(F_1) + (1-\alpha_1)V(G).$$

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

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

So choose $\pi \in \partial V\left(\alpha_1 F_1 + (1-\alpha_1)G\right)$: then

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

Replacing $V(F_1)$ with $\frac{1}{\alpha_1}V\left(\alpha_1 F_1 + (1-\alpha_1)G\right) - \frac{(1-\alpha_1)}{\alpha_1}V(G)$ and, rearranging, we get

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

Multiplying both sides by $\frac{\alpha_1}{1-\alpha_1}$, adding $\langle \varphi_G, \pi\rangle$ to both sides, and rearranging we get

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

which implies that $\langle \varphi_G, \pi\rangle - V(G) \geq c(\pi)$. Hence, it must be that $\pi \in \partial V(G)$. The analogous argument shows that $\pi \in \partial V(F_1)$, so that $\partial V\left(\alpha_1 F_1 + (1-\alpha_1)G\right) \subset \partial V(F) \cap \partial V(G)$. Since $\alpha_2 V(F_2) + \ldots + \alpha_n V(F_n) = (1-\alpha_1)V(G)$, by the inductive assumption, $\partial V(G) \subset \partial V(F_i)$ for all $i = 2, \ldots, n$. We conclude that $\partial V(\alpha_1 F_1 + \ldots + \alpha_n F_n) \subset \partial V(F_i)$ for all $i = 1, \ldots, n$.

Since $\partial V(\alpha_1 F_1 + \ldots + \alpha_n F_n)$ is non-empty, (ii) implies (iii).

To see that (iii) implies (i), choose some $\pi \in \partial V(F_1) \cap \ldots \cap \partial V(F_n)$. Then

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

By Jensen's inequality, convexity of V implies $\alpha_1 V(F_1) + \ldots + \alpha_n V(F_n) \geq V\left(\alpha_1 F_1 + \ldots + \alpha_n F_n\right)$, and therefore (i) holds.

Proof of Proposition 3

The equivalence of (i) and (ii) in the proposition follows directly from the equivalence of (i) and (iii) in Lemma 2.

Proof of Proposition 4

To prove that (i) implies (ii), assume that DM1 is more attentive than DM2 and $(u_1, \bar{p}_1) = (u_2, \bar{p}_2)$ (without loss of generality). 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

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

An envelope theorem (Milgrom and Segal [2002, Theorem 2]) guarantees that

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

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

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

Since DM1 is more attentive 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

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

By monotonicity of the integral,

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

Therefore,

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

Now, to prove that (ii) implies (i), assume that for each pair of menu 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. 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

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

Notice that (i) V_i is a niveloid, hence continuous, and (ii) $\partial V_i(\varphi_F)$ is the subdifferential 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 $\alpha \in [0,1]$,

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

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

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

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

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

Therefore,

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

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 $\eta > 0$,

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

On the other hand, 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

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

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

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

□

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

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

Using translation invariance we can rewrite the above statement as

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

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,

$$\max_{\pi_1 \in \partial V_1(\varphi_F)} \langle \varphi, \pi_1 \rangle = \inf_{\epsilon>0} \frac{V_1(\varphi_F + \epsilon\varphi) - V_1(\varphi_F)}{\epsilon} \geq \inf_{\epsilon>0} \frac{V_2(\varphi_F + \epsilon\varphi) - V_2(\varphi_F)}{\epsilon} = \max_{\pi_2 \in \partial V_2(\varphi_F)} \langle \varphi, \pi_2 \rangle.$$

The next claim completes the proof.

Claim 15. *Fix menu F and assume that $\max_{\pi_1 \in \partial V_1(\varphi_F)} \langle \varphi, \pi_1 \rangle \geq \max_{\pi_2 \in \partial V_2(\varphi_F)} \langle \varphi, \pi_2 \rangle$ for all $\varphi \in \Phi$. Then for all $\pi_2 \in \partial V_2(\varphi_F)$ there is $\pi_1 \in \partial V_1(\varphi_F)$ such that $\pi_1 \trianglerighteq \pi_2$.*

Proof. We show the contrapositive. Suppose that for some $\pi_2 \in \partial V_2(\varphi_F)$ there is no $\pi_1 \in \partial V_1(\varphi_F)$ such that $\pi_1 \trianglerighteq \pi_2$, that is, suppose that $\partial V_1(\varphi_F)$ is disjoint from $\pi_2 + \Phi^*$. Since $\partial V_1(\varphi_F)$ 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 $\varphi \in C(\Delta(\Omega))$ such that, for every $\pi \in \Phi^*$ and $\pi_1 \in \partial V_1(\varphi_F)$, $\langle \varphi, \pi_2 + \pi \rangle > \langle \varphi, \pi_1 \rangle$. 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 $\langle \varphi, \pi \rangle \geqslant 0$ for every $\pi \in \Phi^*$. Since $\Phi^{**} = \Phi$, we have that $\varphi \in \Phi$ (see Section A.1). Since $0 \in \Phi^*$, $\langle \varphi, \pi_2 \rangle > \langle \varphi, \pi_1 \rangle$ for every $\pi_1 \in \partial V_1(\varphi_F)$. □

Proof of Proposition 5

It is straightforward to prove that a constrained attention preference $\succsim$ satisfies Axiom 7, and the intuition for this result is given in Section 4.5. To prove the converse implication, let $\succsim$ be a rationally inattentive preference represented by $(u, \bar{p}, c)$, where c is canonical. Assume without loss of generality that $0 \in u(X)$. We want to show that if $\succsim$ satisfies Axiom 7, then it is a constrained attention preference.

So assume that $\succsim$ satisfies Axiom 7. Define $V : \Phi_{\mathbb{F}} \to \mathbb{R}$ such that for each menu F, $V(\varphi_F) = \max_{\pi \in \Pi(\bar{p})} \langle \varphi_F, \pi \rangle - c(\pi)$, and, by Theorem 2, $c(\pi) = \sup_{F \in \mathbb{F}} \langle \varphi_F, \pi \rangle - V(\varphi_F)$ for every $\pi \in \Pi(\bar{p})$.

We first show that Axiom 7 implies that $V(\alpha\varphi_F) = \alpha V(\varphi_F)$ for all $F \in \mathbb{F}$ with $\varphi_F \geq 0$, and $\alpha > 0$. (The assumption $\varphi_F \geq 0$ guarantees that $\alpha\varphi_F \in \Phi_{\mathbb{F}}$ for all $\alpha > 0$). Fix a menu F such that $\varphi_F \geq 0$. First choose some$\alpha \in (0,1)$. Choose an act $f \sim F$, and an act g such that $\varphi_g = 0$. By Axiom 7, $\alpha F + (1-\alpha)g \sim \alpha f + (1-\alpha)g$, and so

$$V(\alpha\varphi_F) = V(\alpha\varphi_f) = \alpha\varphi_f(\bar{p}) = \alpha V(\varphi_f) = \alpha V(\varphi_F).$$

The case $\alpha = 1$ is obvious. Now assume that $\alpha > 1$. Then,

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

Finally, we show that $c(\pi) > 0$ implies $c(\pi) = \infty$. If so, then we can define $\Gamma = \{\pi : c(\pi) = 0\}$ and complete the proof. Notice that if $c(\pi) > 0$, then there exists a menu F such that $\langle\varphi_F, \pi\rangle - V(\varphi_F) > 0$. By translation invariance, without loss of generality $\varphi_F \geq 0$. Then,

$$\infty \geq c(\pi) \geq \sup_{n\in\mathbb{N}} \langle n\varphi_F, \pi\rangle - V(n\varphi_F) = \sup_{n\in\mathbb{N}} n\left(\langle\varphi_F, \pi\rangle - V(\varphi_F)\right) = \infty.$$

Proof of Proposition 6

It is straightforward to prove that a binary attention preference $\succsim$ satisfies Axiom 8, and the intuition for this result is given in Section 4.5. To prove the converse implication, let $\succsim$ be a rationally inattentive preference $(u, \bar{p}, c)$, where c is canonical. We want to show that if $\succsim$ satisfies Axiom 8, then it is a binary attention preference.

So assume that $\succsim$ satisfies Axiom 8. It is always true that $F \succsim f$ for all $f \in F$, so that we can partition $\mathbb{F}$ into two sets:

$$\mathbb{F}_0 = \{F \in \mathbb{F} : F \sim f \text{ for some } f \in F\} \text{ and } \mathbb{F}_1 = \{F \in \mathbb{F} : F \succ f \text{ for all } f \in F\}.$$

Since $\pi_0 \in \partial V(f)$ for all acts $f \in \mathcal{F}$, notice that $\mathbb{F}_0 = \{F \in \mathbb{F} : \pi_0 \in \partial V(F)\}$. If there is some $\pi_1 \in \bigcap_{F\in\mathbb{F}_1} \partial V(F)$, then $\succsim$ is a binary attention preference.

Claim 16. Let $F_1, \ldots, F_n \in \mathbb{F}_1$. Then $\partial V(F_1) \cap \ldots \cap \partial V(F_n) \neq \varnothing$.

Proof. We prove the claim by induction on n. If $n = 1$, the claim is trivially true. So assume that $\partial V(F_1) \cap \ldots \cap \partial V(F_{n-1}) \neq \varnothing$. By Lemma 2 we know that

$$\partial V(\frac{1}{n-1}F_1 + \ldots + \frac{1}{n-1}F_{n-1}) \subset \partial V(F_1) \quad \Rightarrow \quad \frac{1}{n-1}F_1 + \ldots + \frac{1}{n-1}F_{n-1} \in \mathbb{F}_1,$$

where the last implication holds because $\pi_0 \notin \partial V(F_1)$. Applying the contrapositive of Axiom 8 to F_n and $\frac{1}{n-1}F_1 + \ldots + \frac{1}{n-1}F_{n-1}$ we see that

$$\frac{1}{n}V(F_n) + \frac{n-1}{n}V(\frac{1}{n-1}F_1 + \ldots + \frac{1}{n-1}F_{n-1}) = V(\frac{1}{n}F_1 + \ldots + \frac{1}{n}F_n),$$

which implies by Lemma 2 that

$$\partial V(F_n) \cap \partial V(\frac{1}{n-1}F_1 + \ldots + \frac{1}{n-1}F_{n-1}) \neq \varnothing.$$

Hence $\partial V(F_1) \cap \ldots \cap \partial V(F_n) \neq \varnothing$. □

Since $\Pi(\bar{p})$ is compact, and the collection of closed sets $\{\partial V(F) : F \in \mathbb{F}_1\}$ has the finite intersection property (Claim 16), we conclude that $\bigcap_{F\in\mathbb{F}_1} \partial V(F) \neq \varnothing$.

Proof of Proposition 7

It is straightforward to prove that an exogenous attention preference $\succsim$ satisfies Axiom 9, and the intuition for this result is given in Section 4.5. To prove the converse implication, let $\succsim$ be a rationally inattentive preference $(u, \bar{p}, c)$. We want to show that if $\succsim$ satisfies Axiom 9, then it is an exogenous attention preference.

By Axiom 9, for all menus F and G and $\alpha \in (0, 1)$, $\alpha V(F) + (1-\alpha)V(G) = V(\alpha F + (1-\alpha)G)$. By induction it is easy to see that for all menus $F_1, \ldots, F_n$, and $\alpha_1, \ldots, \alpha_n > 0$ summing up to one,

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

By Lemma 2, $\partial V(F_1) \cap \ldots \cap \partial V(F_n) \neq \varnothing$. Hence, the collection of closed sets $\{\partial V(F) : F \in \mathbb{F}\}$ has the finite intersection property. Since $\Pi(\bar{p})$ is compact, we conclude that $\bigcap_{F\in\mathbb{F}} \partial V(F) \neq \varnothing$. Hence, we can choose some $\pi \in \bigcap_{F\in\mathbb{F}} \partial V(F)$ 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 an exogenous attention preference.

A.3 Solving the consumption-saving problem

In this section we sketch the solution for the consumption-saving example in Section 3.2. The problem is to find x_1, S, and x_2 (conditional on the realization s of signal S) to maximize

$$v(x_1) + E[v(x_2(S)) + v(W - x_1 - x_2(S))] - c(S), \tag{3}$$

with the following parametric restrictions: (i) $v(x) = -e^{-x}$ for $x \in \mathbb{R}$, (ii) W and S are jointly standard normal with correlation ρ, and (iii) $c(S) = -\lambda \log(\sqrt{1-\rho^2})$ where $\lambda > 0$.

Given first period consumption x_1 and signal realization s, the agent chooses x_2 to maximize

$$E[v(x_2) + v(W - x_1 - x_2) \,|\, S = s],$$

leading to the policy function

$$x_2(s) = \frac{1}{2}(-\log(E[e^{-W} \mid S = s]) - x_1).$$

Since W and S are jointly normal, the conditional distribution of W given $S = s$ is normal with $E[W \mid S = s] = \rho s$ and $Var[W \mid S = s] = 1 - \rho^2$. Using the moment-generating function we obtain $E[e^{-W} \mid S = s] = -\rho s + \frac{1}{2}(1 - \rho^2)$, and so

$$x_2(s) = \frac{1}{2}\left((\rho s - x_1) - \frac{1}{2}(1 - \rho^2)\right).$$

The optimal x_1 and S then solve the following problem:

$$v(x_1) + E\left[v\left(\frac{1}{2}(\rho S - x_1) - \frac{1}{4}(1 - \rho^2)\right) + v\left(W - \frac{1}{2}(\rho S + x_1) + \frac{1}{4}(1 - \rho^2)\right)\right] - c(S). \tag{4}$$

Using the functional forms of v and $c(S)$, and the moment-generating function for linear combinations of normal random variables, the objective function becomes

$$-e^{-x_1} - 2e^{\frac{1}{2}x_1 + \frac{1}{4} - \frac{1}{8}\rho^2} + \lambda \log(\sqrt{1 - \rho^2}), \tag{5}$$

Since the signal affects the objective only through ρ^2 (and $\rho^2 = 1$ is never optimal because $\log(0) = -\infty$), it is possible to find solutions for x_1 and S from the following optimality conditions of (5) with respect to x_1 and ρ^2:

$$2\lambda - (1 - \rho^2)e^{\frac{1}{2}x_1 + \frac{1}{4} - \frac{1}{8}\rho^2} \geq 0, \ \rho^2 \geq 0, \ \rho^2\left[2\lambda - (1 - \rho^2)e^{\frac{1}{2}x_1 + \frac{1}{4} - \frac{1}{8}\rho^2}\right] = 0, \text{ and } x_1 = -\frac{2 - \rho^2}{12}.$$

As a result, $\rho^2 = 0$ and $x_1 = -1/6$ are optimal when $e^{\frac{1}{6}} \leq 2\lambda$, and otherwise the optimal ρ^2 and x_1 solve the following equations

$$(1 - \rho^2)e^{\frac{1}{2}x_1 + \frac{1}{4} - \frac{1}{8}\rho^2} = 2\lambda \quad \text{and} \quad x_1 = -\frac{2 - \rho^2}{12}.$$

A.4 Desire for early resolution of uncertainty

To support the interpretation of Axiom 6 as a desire for early resolution of uncertainty, consider an extension of the DM's preferences to the set of lotteries over $\mathbb{F}$ (as, for example, Ergin and Sarver [2010]). Denote by $\alpha \circ F + (1 - \alpha) \circ G$ a lottery over menus that yields F with probability α and G with probability $(1 - \alpha)$, with the interpretation that the lottery realizes *before* the state ω. The DM still chooses acts $f \in F$ and $g \in G$, obtaining outcome $f(\omega)$ or $g(\omega)$ depending on the realization of the lottery and the state ω. However, unlike the mixed menu $\alpha F + (1 - \alpha)G$, the DM can allocate (possibly costly) attention to the realization of the lottery $\alpha \circ F + (1 - \alpha) \circ G$ before she processes information about the state ω.

Now suppose the DM is indifferent between menus F and G. Facing the mixture $\alpha \circ F + (1-\alpha) \circ G$, the DM has the option to acquire information about the mixture, which could reduce uncertainty about which menu is relevant for her payoffs, but she may also incur an attention cost for doing so. If she chooses not to acquire information about the mixture, the menus $\alpha F + (1-\alpha)G$ and $\alpha \circ F + (1-\alpha) \circ G$ are essentially equivalent, and therefore she weakly prefers the "early" realization of the mixture. On the other hand, to avoid the attention costs for mixtures, the DM would prefer to be given menu F or G directly, without the additional uncertainty in $\alpha \circ F + (1-\alpha) \circ G$. The DM therefore weakly prefers F to $\alpha \circ F + (1-\alpha) \circ G$ (because acquiring information about the lottery could be costly), and weakly prefers $\alpha \circ F + (1-\alpha) \circ G$ to $\alpha F + (1-\alpha)G$ (because acquiring information about the lottery could be beneficial). Axiom 6 summarizes these intuitions, reflecting the desire for early resolution of uncertainty.

It is possible also to illustrate this intuition in the context of the consumption-saving example from Section 3.2. Suppose we relax the assumption that income is uncertain only in period 2 (i.e., income stream $(0, W_2, 0)$), and compare the agent's welfare for two alternative income streams, $(0, W_2, W_3)$ or $(W_1, W_2, 0)$, where W_1, W_2 and W_3 are independent and identically distributed. Comparing the two scenarios, the distribution of total wealth is the same, $(W_1 + W_2) \sim (W_2 + W_3)$, and only the timing of realizations is different. In particular, following the menu-choice interpretation offered in Section 3.2, the consumption-saving problem defined by $(0, W_2, W_3)$ can be viewed as a randomization over menus which realize "late" (as in $\alpha F + (1-\alpha)G$), while $(W_1, W_2, 0)$ represents a randomization over menus which realize "early" (as in $\alpha \circ F + (1-\alpha) \circ G$).

First consider the income stream $(0, W_2, W_3)$. In the last period, information about W_3 becomes available to which the agent can pay attention, but such information has no decision-making benefits because consumption in the third period ($x_3 = W_2 + W_3 - x_1 - x_2$) is determined by past decisions (and income realizations). On the other hand, when the income stream is $(W_1, W_2, 0)$, the agent can disregard information about first-period income, obtaining the same welfare as with $(0, W_2, W_3)$. However, she may have a strict incentive to choose an informative signal about W_1, as reducing this uncertainty allows her to make better choices for consumption in periods 1 and 2, and the signal in period 2. Hence, if the cost of extracting an informative signal in period 1 is not too high, the agent strictly prefers the early realization $(W_1, W_2, 0)$. While the opportunity to make better consumption choices with an early realization is present also when there are no attention limitations (e.g., Dreze and Modigliani [1972]), the opportunity to better allocate attention to information about second period income is particular to rational inattention.