UMR 8184

Thema

THéorie Économique, Modélisation et Applications

THEMA Working Paper n°2015-01
Université de Cergy-Pontoise, France

Costly information acquisition and the temporal resolution of uncertainty

Daniele Pennesi

UNIVERSITÉ de Cergy-Pontoise

CNRS CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

Janvier 2015

Costly information acquisition and the temporal resolution of uncertainty

Daniele Pennesi*

January 26, 2015

Abstract

This paper studies the choice of an individual who acquires information before choosing an action from a set of actions, whose consequences depend on the realization of a state of nature. Information processing can be costly, for example, due to limited attention. We show that the preference of the individual over sets of actions, is completely characterized by a preference for early resolution of uncertainty who becomes indifference when facing degenerate choices. When information acquisition is no longer part of the decision process, the individual is indifferent to the timing of resolution of uncertainty and she behaves according to the subjective learning model of Dillenberger et al. (2014).

1 Introduction

Information, uncertainty and time are the essential dimensions of most economic decisions. Typically, information is used to reduce uncertainty and uncertainty resolves over time. In many situations, however, the acquisition of information is costly: for example, when there are bounds to the computational ability of an individual or when information has a market price (for example a data set). The present paper studies the interaction of time, uncertainty and costly information acquisition.

Consider an American tourist who is willing to visit one European country, either Greece (G) or Italy (I). They are "menus of actions". For example the actions in G may be {Kos, Athens} and the actions in I may be {Florence, Venice, Rome}. The payoff of each action depends on the state of nature: the weather. Whereas, the choice of the destination, G or I, depends on the exchange rate between the dollar and the euro.[1] The tourist wants to

*THÉMA (University of Cergy-Pontoise). Email. *daniele.pennesi@u-cergy.fr*

[1] For example, if the rate is higher than a given threshold she will choose G, otherwise she will choose I. Let assume that, with probability α, the rate is higher than the threshold.

acquire information about the weather forecasts to select the best option from each menu and she may allocate attention to different web's sites.[2] She can do it in two different ways

1. Before observing the exchange rate

2. After observing the exchange rate

If she does it before, she will acquire information concerning the weather forecasts of both Greece and Italy and she will schedule her trip contingently to the exchange rate. If she observes the exchange rate before acquiring information, she will focus only to the weather forecast that is relevant to the location she will visit. It is reasonable to imagine that, an individual who pays a cost to acquire information, will prefer to observe the exchange rate before bearing that cost. The result of the paper establishes that this is *the only consideration of an individual who pays a cost to acquire information.* In other words, we interpret uncertainty who resolves before information acquisition (as in 1.), as early resolution of uncertainty. Whereas, late resolution corresponds to a randomization device (the exchange rate) that resolves its uncertainty after information is acquired. We show that the preference over menus of actions of an individual who acquires information at a given cost, is characterized by a preference for early resolution of uncertainty, who becomes indifference when the decision is degenerate. This last consideration follow from the interpretation of the information acquisition problem: information is instrumental to perform a better choice *from* the menu, when the menu contains a single action there is no choice to make.

In addition, when the individual is indifferent to the timing of resolution of uncertainty, i.e. indifference between observing the exchange rate before or after scheduling her vacation, her preferences are represented by the Subjective Learning model of Dillenberger et al. (2014). She will acquire information from a unique source.

With respect to the standing literature, the current paper identifies a definition of preference for early resolution of uncertainty, that is necessary and sufficient to characterize Costly Information Acquisition (CIA). Indeed, as shown in Ex. 1, the definition of early resolution of uncertainty proposed in the current literature is too weak to characterize CIA.

The results of the paper raise questions concerning observational distinguishability of a model of costly information acquisition from a pure preference for the timing of resolution

[2] Searching for "weather forecast Italy" and "weather forecast Greece" on the web gives, respectively, 10,200,000 and 6,500,000 results.

of uncertainty. As pointed out above, information is instrumental to select the best action from a menu of actions, when facing degenerate menus i.e. singletons, costly information acquisition is not effective. Therefore, the timing of resolution of uncertainty for degenerate menus is irrelevant, this implies that CIA is distinguishable from a pure preference for early resolution of uncertainty and this distinction can be used to discriminate experimentally the CIA model from those valuing early resolution of uncertainty intrinsically.

Concerning the Subjective Learning model, there is no observational distinguishability from a pure indifference toward the timing of resolution of uncertainty.

The paper also offers a new rationale for preference for early resolution of uncertainty. Whenever an individual displays a preference for early resolution of uncertainty over non-degenerate menus, she behaves "as if", she solves an optimal information acquisition problem with costly information. Therefore, the behavioral foundation of a preference for early resolution of uncertainty, may be traced back to an aversion to contingent planning that follows from the cost of acquiring information.

2 Overview of the results

An individual with CIA preferences, optimally chooses an action (acts) f from a set of possible actions F. Each action associates a payoff $f(\omega) \in X$ to the state of the world $\omega \in \Omega$ that will realize. Information about the true state of the world is acquired through a *channel* (or experiment). The channel specifies the probability of forming a posterior p given a prior $\hat{p}$. After information is received, the decision maker selects an act from the menu. Formally:

$$\max_{\pi \in \Pi(\hat{p})} \left[\int_{\Delta(\Omega)} \max_{f \in F} \left(\int_{\Omega} u(f(\omega))\, p(d\omega) \right) \pi(dp) - c(\pi) \right] \tag{CIA}$$

where $u : X \to \mathbb{R}$ is a utility over the payoffs in X, $\Pi(\hat{p})$ is the set of all channels for a given prior $\hat{p}$ and $c : \Pi(\hat{p}) \to [0,\infty]$ is the cost of information.

The CIA model has been axiomatized by De Oliveira et al. (2013) and it is interpreted as representing the preferences of a Rationally Inattentive individual (in the sense of Sims, 2003). The timing of the rational inattentive choice procedure is the following: an unobservable (to the decision maker) state of the world is selected by nature, the decision maker chooses a menu from which she will select an act later. After choosing a menu, information is acquired and the posterior is formed. Lastly, the individual selects an act from the menu and she receives the payoff associated with the state of the world and the selected act. The decision maker may also face randomized menus (as in the exchange rate example), $\alpha F + (1-\alpha)G$, where a biased coin is tossed and a menu is selected according to the landing of the coin. For randomized menus there are two layers of uncertainty: the state of the world and the realization of the biased coin. In the axiomatization of De Oliveira et al. (2013), the realization of the biased coin is known to the individual only *after information is acquired* (Pag 6). Hence, the individual always sets up a plan of actions contingent to the realization of the biased coin. In the present work, we assume a more general domain

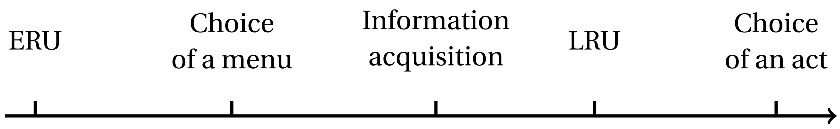

Figure 1: Early (ERU) and late (LRU) resolution of uncertainty

in which revealed preferences are observed, that allows to distinguish between later and earlier resolution of uncertainty. Indeed, our primitive is a preference over lotteries over menus of acts. We allow for a different form of randomization between menus that resolves its uncertainty *before* information is acquired. Figure 1 illustrates the timing of the decision process in the two cases. LRU is the point in time at which randomization takes place in De Oliveira et al. (2013), we allow an additional form of randomization that resolves its uncertainty at ERU.

As in Ergin and Sarver (2014), the early randomization is given by the lottery $\alpha\delta_F + (1-\alpha)\delta_G$, it is a lottery that pays F with probability α and G with probability $(1-\alpha)$ and it is played immediately. It corresponds to the left-hand side of Fig. 2. In this case, the individual knows the menu from which she will choose in the second stage, before acquiring information.

When facing $\delta_{\alpha F+(1-\alpha)G}$ the individual acquires information to perform contingent choices.

It is easy to see that a CIA individual will prefer the left-hand side to the right-hand side of Fig. 2. This is due to the cost of acquiring information, the choice in the right-hand side of Figure 2, is "more complex" in terms of the amount of information to acquire, hence more costly. The opposite implication, that preferring the left-hand side to the right-hand side of Figure 2 implies CIA is, however, not true. As claimed above, when facing random-

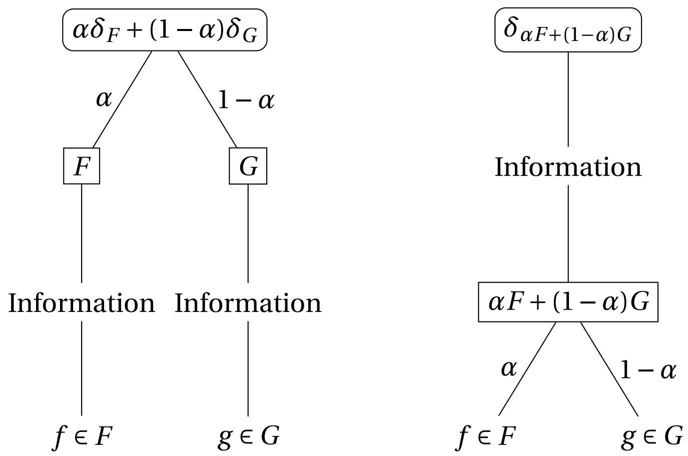

Figure 2: Early resolution of uncertainty (left) and late resolution of uncertainty (right).

ization between degenerate menus, the acquisition of information becomes superfluous, since information is used to perform a better choice from the menu. Therefore, denoting f the degenerate menu $\{f\}$, the CIA model implies $\alpha\delta_f + (1-\alpha)\delta_g \sim \delta_{\alpha f+(1-\alpha)g}$. This is not necessarily true for a pure preference for early resolution of uncertainty: in that case, an intrinsic preference for early resolution of uncertainty may follow from reasons such as, anxiety or the possibility to take hidden actions (see Ergin and Sarver (2014)), hence a strict preference $\alpha\delta_f + (1-\alpha)\delta_g \succ \delta_{\alpha f+(1-\alpha)g}$, is possible. Theorem 1 establishes that the CIA model is completely characterized by the existence of a strict preference for early resolution of uncertainty when non-degenerate menus are taken into account, i.e. for some $F, G \in \mathscr{A}$ with $|F| > 1$ and $|G| > 1$ and $\alpha \in (0,1)$, $\alpha\delta_F + (1-\alpha)\delta_G \succ \delta_{\alpha F+(1-\alpha)G}$.

When indifference toward the timing of resolution of uncertainty is extended to all menus, i.e. $\alpha\delta_F + (1-\alpha)\delta_G \sim \delta_{\alpha F+(1-\alpha)G}$, preferences are represented by a particular case of CIA, namely the Subjective Learning (SL) model of Dillenberger et al. (2014). In SL, the decision maker acquires information from a unique source, therefore making it before or after the randomization is performed, is equivalent. In the initial example, the American tourist is indifferent between knowing the exchange rate before or after acquiring information, if and only if, she gathers information concerning the weather forecasts of a single country, for example Greece.

As a final result, we prove that the attitude toward the timing of resolution of uncertainty is related to the solution of the optimal information acquisition problem. Suppose there exists a source of information that is optimal for two different menus, i.e. a π^* that solves the maximization in Eq. (CIA) for two menus F and G. If this is the case, the individual is indifferent between acquiring information before or after the randomization between F and G, i.e. $\alpha\delta_F + (1-\alpha)\delta_G \sim \delta_{\alpha F+(1-\alpha)G}$. In other words, she "locally" behave according to the SL model.

2.1 Related literature

The acquisition of information may be costly for a variety of reasons. One of these, limited attention, has been recently studied by different authors. It has introduced with a series of papers by C. Sims (Sims, 2003, 1998) as a natural explanation for price stickiness. The model has been subsequently applied to a variety of topics.[3]

The behavioral foundation of the costly information acquisition due to inattention has been studied over the last years. Recently, Masatlioglu et al. (2012), Caplin and Dean (2014) and Ellis (2013) characterize rational inattentive preferences observing *ex post* choices from menus of alternatives. In the current paper, we use choice over menus of actions. The closest paper to the current one is De Oliveira et al. (2013). They use choice over menus of actions to axiomatize the model in Eq. (CIA), in which an individual acquires costly information (pays attention) to optimally select an action from a menu of actions. The focus of the current paper is to provide an alternative behavioral foundation of the same model, that relies on the distinction between early and late resolution of uncertainty. A full comparison with De Oliveira et al. (2013) is discussed in Section 3.2. de Oliveira (2014) characterizes a particular case of the CIA model, with an entropic cost of attention who corresponds to the formulation of the RI model originally proposed by Sims (2003).

Dillenberger et al. (2014) axiomatized another particular case of the CIA model, called Subjective Learning (SL) model in which attention is allocated to a unique source of information. This paper modifies the approach of De Oliveira et al. (2013) and Dillenberger et al. (2014), introducing a distinction between early and late resolution of uncertainty, in the spirit of Kreps and Porteus (1978). The current paper is also related to the work of Ergin and Sarver (2014). They study intrinsic preference for early resolution of uncertainty when choices are observed over lotteries over menus of lotteries. Therefore, the two settings are clearly different. Moreover, (second-stage) uncertainty in their model is subjec-

[3] Price setting (Mackowiak and Wiederholt, 2009), optimal consumption/saving (Luo, 2008; Tutino, 2008), business cycles (Kacperczyk et al., 2009), portfolio under-diversification (Van Nieuwerburgh and Veldkamp, 2009), asset pricing (Mondria, 2010), stochastic choice (Matějka and McKay, 2014).

tive, whereas in our model is objective i.e. is represented by Ω. A full comparison with their paper is contained in Section 3.3.

3 Axiomatic foundation

A finite set Ω contains the states of the world, when a state is realized, all uncertainty is resolved. We denote X the set of consequences, it is a convex subset of $\mathbb{R}$. An act is a function $f : \Omega \to X$, the set of all acts is denoted by $\mathscr{F}$. A constant act $f(\omega) = x$ for all $\omega \in \Omega$ is identified with an element of X. We denote $\mathscr{A}$ the set[4] of all nonempty and finite subsets of $\mathscr{F}$. Given a set M, we denote $\Delta(M)$ the set of probabilities defined on M. The preference relation $\succcurlyeq$ is defined over lotteries over menus of acts i.e. on $\Delta(\mathscr{A})$. We denote f the degenerate menu $F = \{f\}$. x denotes the degenerate menu containing only a constant act $f(\omega) = x$ for all $\omega \in \Omega$ and some $x \in X$. δ_F denote the degenerate lottery paying $F \in \mathscr{A}$ for sure. The mixture + of two acts $\alpha f + (1-\alpha)g$ for all $\alpha \in [0,1]$ is performed state-wise, $(\alpha f + (1-\alpha)g)(\omega) = \alpha f(\omega) + (1-\alpha)g(\omega)$. The mixture of two menus $\alpha F + (1-\alpha)G$ for all $\alpha \in [0,1]$ is the menu of all mixtures of elements in F and G, $\alpha F + (1-\alpha)G = \{\alpha f + (1-\alpha)g : f \in F,\ g \in G\}$.

The first axiom includes some standard properties whose interpretation is well established.

Axiom (Preference)**.**

1. (Weak Order) *$\succcurlyeq$ is a non-trivial weak order.*
2. (Continuity) *The upper and lower contour sets, $\{P \in \Delta(\mathscr{A}) : P \succcurlyeq Q\}$ and $\{P \in \Delta(\mathscr{A}) : P \preccurlyeq Q\}$ are closed in the weak* topology.*
3. (Dominance) *Given $\delta_F, \delta_G \in \Delta(\mathscr{A})$, if for all $g \in G$ there exists $f \in F$ with $\delta_{f(\omega)} \succcurlyeq \delta_{g(\omega)}$ for each $\omega \in \Omega$, then $\delta_F \succcurlyeq \delta_G$.*
4. (Unboundedness) *There are $x, y \in X$ such that, for each $\alpha \in (0,1)$, there exists $z, z' \in X$ such that $\delta_{\alpha z + (1-\alpha)y} \succ \delta_x \succ \delta_{\alpha z' + (1-\alpha)y}$.*

Axioms 1. and 2. are minimal rationality and continuity requirements. Axiom 3. is introduced (in their setting) in De Oliveira et al. (2013), it implies a preference for flexibility, i.e. if $G \subseteq F$ then $\delta_F \succcurlyeq \delta_G$, and the monotonicity axiom used in choice under ambiguity i.e.

[4] We identify $\mathscr{A}$ as the power set of X^S. Given the Euclidean metric d, we endow $\mathscr{A}$ with the Hausdorff distance given by $d_h(F,G) = \max\left\{\sup_{f \in F} \inf_{g \in G} d(f,g), \sup_{f \in G} \inf_{g \in F} d(f,g)\right\}$.

if $\delta_{f(\omega)} \succcurlyeq \delta_{g(\omega)}$ for all $\omega \in \Omega$ then $\delta_f \succcurlyeq \delta_g$. Axiom 4. implies that $u(X) = \mathbb{R}$, it only restricts preferences over final prizes.

Next axiom is the classical independence axiom with respect to lotteries.

Axiom (Ex-ante Independence). *For any $P, Q, T \in \Delta(\mathscr{A})$ and $\alpha \in (0,1)$,*

$$P \succcurlyeq Q \iff \alpha P + (1-\alpha)R \succcurlyeq \alpha Q + (1-\alpha)R$$

The interpretation is standard, a preference for a lottery P over Q is not reversed when mixing with a third lottery.

The next definition formally introduces the preference for early resolution of uncertainty (PERU), as in Ergin and Sarver (2014).

(PERU). *For any $F, G \in \mathscr{A}$ and $\alpha \in (0,1)$,*

$$\alpha\delta_F + (1-\alpha)\delta_G \succcurlyeq \delta_{\alpha F + (1-\alpha)G}$$

It implies that the left "tree" of Figure 2 is weakly preferred to the right one.

Similarly, we define indifference toward the timing of resolution of uncertainty:

(ITRU). *For any $F, G \in \mathscr{A}$ and $\alpha \in (0,1)$,*

$$\alpha\delta_F + (1-\alpha)\delta_G \sim \delta_{\alpha F + (1-\alpha)G}$$

Next property imposes ITRU only with respect to lotteries involving degenerate menus, we name it Degenerate ITRU, more precisely:

(DITRU). *For any $f, g \in \mathscr{F}$ and $\alpha \in (0,1)$,*

$$\alpha\delta_f + (1-\alpha)\delta_g \sim \delta_{\alpha f + (1-\alpha)g}$$

Before stating the main theorem, we formally introduce the Costly Information Acquisition and the Subjective Learning representations of preferences. First, we need some preliminary definitions:

Definition 1. *Given a prior* $\hat{p} \in \Delta(\Omega)$, *a* channel *is a probability* $\pi \in \Delta(\Delta(\Omega))$ *such that*

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

A channel gives the probability of obtaining a posterior for a given prior $\hat{p}$. The convex set of all channels relative to a prior $\hat{p}$ is denoted by $\Pi(\hat{p})$. Next definition introduces the information cost function:

Definition 2. *Given a prior* $\hat{p} \in \Delta(\Omega)$, *a function* $c : \Pi(p) \to [0,\infty]$ *is an* information cost function *if it is lower semicontinuous and it satisfies:*

(i) $c(\pi) = 0$ *whenever* $\pi(\hat{p}) = 1$.

(ii) $c(\alpha\pi + (1-\alpha)\rho) \leq \alpha c(\pi) + (1-\alpha)c(\rho)$ *for all* $\pi, \rho \in \Pi(\hat{p})$ *and* $\alpha \in (0,1)$.

(iii) $c(\rho) \leq c(\pi)$ *for all* $\pi, \rho \in \Pi(\hat{p})$ *and* $\int_{\Delta(\Omega)} a(p)\pi(dp) \geq \int_{\Delta(\Omega)} a(p)d\rho(p)$ *for all convex and continuous* $a : \Delta(\Omega) \to \mathbb{R}$.

Property (i) states that acquiring no information is costless. Property (ii) and lower semi-continuity are regularity conditions satisfied by the cost functions used in the literature. Condition (iii) states that more informative channels (in the sense of Blackwell) are more costly.

Definition 3. *A preference* $\succcurlyeq$ *has a Costly Information Acquisition representation if, there exists a tuple* $(V, \Pi(\hat{p}), \hat{p}, u, c)$, *where* $\Pi(\hat{p}) \subset \Delta(\Delta(\Omega))$ *is a set of channels,* $\hat{p} \in \Delta(\Omega))$ *is a prior,* $u : X \to \mathbb{R}$ *is a Bernoulli utility and* $c : \Delta(\Delta(\Omega)) \to [0,\infty]$ *is an information cost function and* $V : \mathscr{A} \to \mathbb{R}$ *is such that, for all* $P, Q \in \Delta(\mathscr{A})$, $P \succcurlyeq Q \iff E_P[V] \geq E_Q[V]$ *and*

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

The Subjective Learning representation of Dillenberger et al. (2014), is a particular case of the CIA representation, in which information is gathered from a unique source. Formally,

Definition 4. *A preference* $\succcurlyeq$ *has a Subjective Learning representation if, there exists a tuple* (V, π, u) *where,* $\pi \in \Delta(\Delta(\Omega))$ *is a channel,* $u : X \to \mathbb{R}$ *is a Bernoulli utility and* $V : \mathscr{A} \to \mathbb{R}$ *is such that, for all* $P, Q \in \Delta(\mathscr{A})$, $P \succcurlyeq Q \iff E_P[V] \geq E_Q[V]$ *and*

$$V(F) = \int_{\Delta(\Omega)} \max_{f\in F} \left(\int_{\Omega} u \circ f \, dp \right) d\pi$$

In the Subjective Learning model, the choice of information is not part of the decision process.

Next theorem is the main result of the paper:

Theorem 1. *Given a preference $\succcurlyeq$ defined over $\mathscr{A}$, then:*

a) $\succcurlyeq$ satisfies Preference, Ex-ante Independence, PERU and DITRU, if and only if, $\succcurlyeq$ has a Costly Information Acquisition representation.

b) $\succcurlyeq$ satisfies Preference, Ex-ante Independence and ITRU, if and only if, $\succcurlyeq$ has a Subjective Learning representation.

Theorem 1 characterizes the relation between attitude toward the timing of resolution of uncertainty and the acquisition of costly information. A preference for early resolution who becomes indifference when restricted to degenerate menus, completely characterizes the costly information acquisition model.[5] Indeed, when mixing singletons, the timing of resolution of uncertainty is ineffective. This is a natural consequence of the information acquisition problem modelled by CIA. When facing two degenerate menus of acts, information is not valuable, since information is instrumental to perform a better choice *from* the menu. For singleton menus there is no gain in acquiring information. As a consequence, costly information acquisition does not play a role in this situation. This is the main difference between the CIA representation and a pure PERU.

When the timing of resolution of uncertainty is not relevant (e.g. ITRU), the preference is represented by the Subjective Learning model. The individual acquires information from the unique channel π.

The indifference toward the timing of resolution of uncertainty for degenerate menus can be also imposed using a weak independence axiom restricted to singletons. It is useful to provide this alternative axiomatization to perform a direct comparison with De Oliveira et al. (2013). Indeed, the next axiom is a weakening of their Weak Singleton Independence who imposes independence for ex post mixtures of menus with singletons (see Section 3.2):

Axiom (Weak Degenerate Independence (WDI))**.** *For any $f, g, h, h' \in \mathscr{F}$ and $\alpha \in (0,1)$,*

$$\delta_{\alpha f+(1-\alpha)h} \succcurlyeq \delta_{\alpha g+(1-\alpha)h} \implies \delta_{\alpha f+(1-\alpha)h'} \succcurlyeq \delta_{\alpha g+(1-\alpha)h'}$$

[5] More precisely, the CIA model is characterize by PERU with, at least, a strict preference.

It is easy to prove that WDI, together with other axioms, implies DITRU. Therefore, we have the following corollary:

Corollary 1. *A preference $\succcurlyeq$ satisfies axioms Preference, Ex-ante Independence, WDI and PERU, if and only if, $\succcurlyeq$ has a Costly Information Acquisition representation.*

It follows from the previous corollary that PERU alone can characterize CIA if we impose a weak independence axiom over degenerate menus. This implies that we *derive* from PERU, the Weak Singleton Independence axiom of De Oliveira et al. (2013) rather than assuming it

3.1 Optimal information and ITRU

In this section we show that the optimal channels associated to a given menu, i.e. the ones solving the optimization problem in (CIA), convey information about the attitude toward the resolution of uncertainty. In particular, if two menus share an optimizing channel, then the individual is indifferent to mixing them before or after acquiring the information. More precisely, consider the set of optimal channels for a given menu $F \in \mathscr{A}$,

$$\partial V(F) = \underset{\pi \in \Pi(\hat{p})}{\operatorname{argmax}} \left[\int_{\Delta(\Omega)} \max_{f \in F} \left(\int_{\Omega} u \circ f \, dp \right) d\pi - c(\pi) \right]$$

and take two menus $F, G \in \mathscr{A}$. If there exists a channel π^* belonging to $\partial V(F)$ and to $\partial V(G)$, it means that the optimal information acquisition related to F and the one related to G have, at least, a common solution. The mere existence of such a channel, is sufficient to impose indifference toward the timing of randomization, when it involves F and G. Intuitively, if π^* is optimal for both F and G, the individual can always acquire π^*, so she is indifferent to the timing of the randomization, as in the SL model. More precisely:

Proposition 1. *Suppose that $\succcurlyeq$ has a CIA representation, then the following are equivalent:*

1. $\partial V(F) \cap \partial V(G) \neq \emptyset$
2. $\alpha\delta_F + (1-\alpha)\delta_G \sim \delta_{\alpha F + (1-\alpha)G}$ *for all* $\alpha \in (0,1)$.

The result *does not* imply that the actual channel used, when evaluating F and G, is exactly π^*, but the existence of a common optimal channel is sufficient to guarantee indifference. After all, the individual can always acquire π^*. The opposite of condition 2. in

Proposition 1, has been interpreted in De Oliveira et al. (2013) as a "reallocation of attention between F and G", that is, to focus on different form of informations depending on their relevance when facing different menus. This can be used to extend their comparative statics analysis. Suppose we can observe the preferences $\succcurlyeq_1$ and $\succcurlyeq_2$ of two individuals and assume V_1 and V_2 are they representation. The following definition can be found in De Oliveira et al. (2013)

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

The following is a comparative notion of preference for early resolution of uncertainty:

Definition 6. *V_1 has a greater preference for early resolution of uncertainty than V_2 if, for all menus F and G and all $\alpha \in (0,1)$,*

$$\alpha\delta_F + (1-\alpha)\delta_G \succ_2 \delta_{\alpha F+(1-\alpha)G} \implies \alpha\delta_F + (1-\alpha)\delta_G \succ_1 \delta_{\alpha F+(1-\alpha)G}$$

The following theorem establishes an equivalence between tendency to relocate attention and strict preference for early resolution of uncertainty.

Theorem 2. *The following are equivalent:*

1. *V_1 has a greater preference for early resolution of uncertainty than V_2.*
2. *V_1 has a greater tendency to reallocate attention than V_2.*

The two conditions are equivalent, indeed, as proved in De Oliveira et al. (2013), a tendency to reallocate attention is related to a strong preference for early resolution of uncertainty (as defined in their condition (ii) of Theorem 3).

3.2 Relation with De Oliveira et al. (2013)

De Oliveira et al. (2013) proposed an axiomatization of the CIA model that is the starting point of the present work. There are, however, some substantial differences. First, we based our axiomatic on the distinction between early and late resolution of ex-ante uncertainty, i.e. the timing of the randomization between menus. Indeed, our setting allows to distinguish between randomizations taking place before or after information is acquired. The main axiom in our representation is PERU and it captures the additional value that early resolution of uncertainty has for a CIA individual. Differently from the current paper, they introduced an axiom, Aversion to Randomization, that captures a preference for "early resolution of uncertainty". It postulates that, if $F \sim G$ then $F \succcurlyeq \alpha F + (1-\alpha)G$, in our setting this would be equal to $\delta_F \sim \delta_G$ implies $\delta_F \succcurlyeq \delta_{\alpha F+(1-\alpha)G}$. When the agent is indifferent between two menus, she may strictly prefer to commit to one menu rather than randomizing. This is a weak form of preference for "early resolution of uncertainty", however, in their setting there is no distinction between early and later resolution of uncertainty: the randomization takes place always after information is acquired (attention is allocated) (De Oliveira et al., 2013, pag. 6). Moreover, the previous axiom alone is not sufficient to characterize the CIA representation, whereas PERU (and DITRU) completely characterizes CIA. For example a multiplicative cost representation of the following type satisfies Aversion to Randomization but it is not a CIA model.

Example 1. *Let $c(\pi) : \Delta(\Delta(\Omega)) \to [0, \infty]$ be an information cost, then:*

$$V(F) = \max_{p \in \Pi(\hat{p})} \frac{\int \max_{f \in F} \left(\int u(f) dp \right) \pi(dp)}{c(\pi)}$$

satisfies Aversion to Randomization[6].

In addition to Aversion to Randomization, they impose the Weak Singleton Independence (WSI) axiom. It postulates that, for all menus F, G, all $\alpha \in (0,1)$ and all acts h, h', $\alpha F + (1-\alpha)h \succcurlyeq \alpha G + (1-\alpha)h \;\Rightarrow\; \alpha F + (1-\alpha)h' \succcurlyeq \alpha G + (1-\alpha)h'$. In our second axiomatization (Corollary 1), we weakened WSI to Weak Degenerate Independence and we use the fact that PERU is stronger than Aversion to Randomization[7] to derive CIA. Therefore, we *derive* Weak Singleton Independence as a byproduct of a preference for early resolution of uncertainty.

To conclude, our axiomatization constitutes an alternative to that in De Oliveira et al. (2013) that poses more weight on the effect that the timing of randomization has on preferences. It is meant to link two different branches of literature and to propose a new interpretation to costly information acquisition and the preference for early resolution of uncertainty.

[6]See Section B.1.

[7]Suppose $F \sim G$, by Ex-ante independence, $\alpha F \oplus (1-\alpha)G \sim G$, by PERU, $G \succcurlyeq \alpha F + (1-\alpha)G$.

3.3 Relation with Ergin and Sarver (2014)

The approach we use in this paper is related to a recent work of Ergin and Sarver (2014). They studied preferences toward the timing of resolution of uncertainty when the objects of choice are lotteries over menus of lotteries. Their Costly Contemplation representation is the closest to the representation of Definition 3 and it ranks lotteries over menus according to $P \mapsto E_P[V]$ where

$$V(F) = \max_{\theta \in \Theta} \int_S \max_{\beta \in F} u(\beta, s; \theta)\pi(ds) - c(\theta)$$

where Θ is interpreted as the set of hidden action a decision maker could take before the resolution of uncertainty. An equivalent formulation of V is the following:

$$V(F) = \max_{\mu \in \mathcal{M}} \int_U \max_{\beta \in F} u(\beta)\mu(u) - c(\mu) \tag{1}$$

where $\mathcal{M}$ is a set of positive *measures*. This second representation is more appropriate to understand the differences with our result. There are two main differences: first, (ex post) uncertainty in Ergin and Sarver (2014) is subjective. It follows from the uncertainty of the individual about her future tastes. Here, we assume the existence of an objective state space Ω and a decision maker who acquires information to form a posterior over Ω. We derive a representation who separates tastes and beliefs, something that is not possible in the setting of Ergin and Sarver (2014). However, the most important difference concerns the hidden action interpretation. They interpret a preference for early resolution of uncertainty as if an unobservable (to the modeller) action can be taken by the individual, before uncertainty is resolved. In the case of the Costly Contemplation, the action is exactly a strategy of contemplation (the observation of a subjective signal about future tastes) of the future uncertainty. When facing a mixture of two non-singleton menus, contemplation is more costly, due to the complexity of the problem faced by the individual. Therefore, the individual prefers early resolution of uncertainty. The hidden action/costly contemplation interpretation follows from a representation of preference, as in Eq. (1), where the set $\mathcal{M}$ contains measures that are not necessarily probabilities. Those measures are therefore converted into "hidden actions". To the contrary, we directly derive a set of channels or *probability* measures over priors $\Pi(\hat{p})$ hence, our representation is free of any additional interpretation, early resolution of uncertainty is valuable only because information is costly and no unobservable hidden action (contemplation strategy) is taken.

4 Conclusion

We axiomatized of a model of Costly Information Acquisition that builds upon the attitude toward the timing of resolution of uncertainty. A preference for early resolution of uncertainty that becomes indifference when facing degenerate menus, completely characterizes the model. Indifference toward the timing of resolution of uncertainty characterizes the Subjective Learning model of Dillenberger et al. (2014). The results provide a method to experimentally distinguish a pure PERU from the CIA model.

A Preliminaries

For completeness, we introduce some notation taken from De Oliveira et al. (2013). $C(\Delta(\Omega))$ is the linear space of continuous real-valued function defined on $\Delta(\Omega)$. $ca(\Delta(\Omega))$ is the linear space of signed measures of bounded variation defined on $\Delta(\Omega)$. $C(\Delta(\Omega))$ is equipped with the supnorm and $ca(\Delta)$ with the weak* topology. $ca(\Delta(\Omega))$ is the continuous dual of $C(\Delta(\Omega))$. We denote Φ the set of convex functions in $C(\Delta(\Omega))$, it is a closed convex cone containing the zero function.

A functional $V : C(\Delta(\Omega)) \to \mathbb{R}$ is said to be normalized if $V(k) = k$ for all $k \in \mathbb{R}$ (where we identify the constant function $1_{\Delta(\Omega)}k$ with k); monotone, if $V(a) \geq V(b)$ whenever $a(p) \succcurlyeq b(p)$ for all $p \in \Delta(\Omega)$; translation invariant if $V(a+k) = V(a)+k$ for all $a \in \mathbb{R}$. A translation invariant functional is (Lipschitz) continuous.

Given an affine utility $u : X \to \mathbb{R}$, we denote $\sigma_F : \Delta(\Omega) \to \mathbb{R}$, the function

$$\sigma_F(p) = \max_{f \in F} \int_\Omega u \circ f \, dp \quad \forall p \in \Delta(\Omega) \tag{2}$$

for some finite menu F. The set of such functions is denoted by $\hat{\Phi}$. If $u(X) \subseteq \mathbb{R}$ is unbounded above, $\hat{\Phi}$ has the following properties (see Lemma 1 De Oliveira et al., 2013) for a proof.

1. $\hat{\Phi} + [0, \infty) = \hat{\Phi}$.
2. $\hat{\Phi} + \mathbb{R}$ is dense in Φ.

B Proofs

By the von Neumann-Morgenstern's theorem, axioms Weak Order, Continuity and Ex-ante Independence are necessary and sufficient to the existence of a continuous function $V : \mathscr{A} \to \mathbb{R}$ such that

$$P \succcurlyeq Q \iff E_P[V] \geq E_Q[V]$$

Proof. Of Theorem 1. Part a). The fact that a preference with a CIA representation implies the axioms is straightforward. To prove the converse implication of Theorem 1 we need a series of preliminary results.

The next lemma introduces a representation of $\succcurlyeq$ over degenerate menus:

Lemma 1. *There exists an affine utility $u : X \to \mathbb{R}$ with unbounded (above and below) range and a probability $\hat{p} \in \Delta(\Omega)$ such that*

$$V(f) = \int_\Omega u(f(\omega))\hat{p}(d\omega)$$

represents $\succcurlyeq$ over $\Delta(\mathscr{F})$. Moreover, $\hat{p}$ is unique and u is unique up to positive affine transformations.

Proof. Proof of Lemma 1. To leverage the Mixture Theorem, we only need to prove independence. Suppose $\delta_f \sim \delta_g$, by Ex-ante Independence, for all $H \in \Delta(\mathscr{A})$ and $\gamma \in (0,1)$, $\gamma\delta_f + (1-\gamma)\delta_H \sim \gamma\delta_g + (1-\gamma)\delta_H$, it clearly holds for all degenerated $H = \{h\}$ in $\mathscr{A}$. By DITRU $\delta_{\gamma f + (1-\gamma)h} \sim \delta_{\gamma g + (1-\gamma)h}$. Hence $\succcurlyeq$ satisfies the independence axiom. An application of the Mixture theorem gives the result. Unboundedness above and below of the function u follows from standard arguments (see Cerreia-Vioglio et al., 2011).

□

We say that the preference $\succcurlyeq$ satisfies Indifference to Randomization if $\delta_F \sim \delta_{\text{co}(F)}$. We begin with a lemma due to Ergin and Sarver (2014, Lemma 5), the proof is identical and we report it for completeness.

Lemma 2. *If a preference $\succcurlyeq$ satisfies Weak Order, Continuity, PERU and Dominance, then it satisfies Indifference to Randomization.*

Proof. Of Lemma 2. By Dominance $\delta_{\text{co}(F)} \succcurlyeq \delta_F$. Now let define recursively $\delta_{F_0} = \delta_F$ and $\delta_{F_k} = \frac{1}{2}\delta_{F_{k-1}} + \frac{1}{2}\delta_{F_{k-1}}$ for $k \geq 1$. By PERU,

$$\delta_{F_{k-1}} = \frac{1}{2}\delta_{F_{k-1}} + \frac{1}{2}\delta_{F_{k-1}} \succcurlyeq \delta_{\frac{1}{2}F_{k-1}+\frac{1}{2}F_{k-1}} = \delta_{F_k}$$

By transitivity $\delta_F \succcurlyeq \delta_{F_k}$ for all k. As $k \to \infty$, $d_h(\delta_{F_k}, \delta_{\text{co}(F)}) \to 0$ and $\delta_{F_k} \to \delta_{\text{co}(F)}$ in the weak* topology, by Continuity, $\delta_F \succcurlyeq \delta_{\text{co}(F)}$. $\square$

Lemma 3. *For each $F \in \mathscr{A}$, there exists a $x_F \in X$ such that $\delta_F \sim \delta_{x_F}$.*

The standard proof follows from Continuity, Dominance and the finiteness of the state space.

Now, let's define functionals $W : \hat{\Phi} \to \mathbb{R}$ as $W(\sigma_F) \triangleq V(x_F)$ with $x_F \in X$ and $\delta_{x_F} \sim \delta_F$ and σ_F is given by Eq. (2). By Lemma 1, $W(\sigma_G) = u(x_F)$. To see that W is well defined, assume $\delta_{x_F} \sim \delta_{y_F}$, then $V(F) = u(x_F) = u(y_F) = V(F)$. Moreover, let assume $\sigma_F = \sigma_G$, we will show that $F \sim G$. First, notice that $\sigma_F = \sigma_G$ implies $\text{co}(u(F)) = \text{co}(u(G))$. Indeed, suppose without loss of generality, the existence of a $m \in \mathbb{R}^{|\Omega|}$ such that $m \in \text{co}(u(F)) \setminus \text{co}(u(G))$. By a separating hyperplane theorem, there exists a $q \in \mathbb{R}^{|\Omega|}$ such that

$$\int_\Omega m dq < \max_{n \in u(G)} \int_\Omega n dq = \max_{g \in G} \int_\Omega u(g(\omega)) q(d\omega)$$

Renormalizing q such that $q \in \Delta(\Omega)$, the last inequality contradicts $\sigma_F = \sigma_G$. By affinity of u, $\text{co}(u(F)) = u(\text{co}(F))$, then $\text{co}(F) = \text{co}(G)$, or equivalently $\text{co}(F) \subseteq \text{co}(G)$ and $\text{co}(G) \subseteq \text{co}(F)$, by Dominance and Indifference to Randomization, $\delta_F \sim \delta_G$. Therefore W is well-defined.

Lemma 4. *$W(\sigma_F)$ is monotone (with respect to point-wise order) and normalized. It is convex if and only if $\succcurlyeq$ satisfies PERU.*

Proof. Of Lemma 4. Monotonicity follows from the previous results. The constant functions in $\hat{\Phi}$ are those taking values in $u(X) = \mathbb{R}$. To see normalization, take $x \in X$, then $W(\sigma_x) = u(x) = \sigma_x$ by definition. To see convexity, for all $\sigma_F, \sigma_G \in \hat{\Phi}$ and all $\alpha \in (0,1)$, PERU implies

$$\begin{aligned}
&\alpha\delta_F + (1-\alpha)\delta_G \succcurlyeq \delta_{\alpha F + (1-\alpha)G} \iff \\
&\alpha V(F) + (1-\alpha)V(G) \geq V(\alpha F + (1-\alpha)G) \iff \\
&\alpha W(\sigma_F) + (1-\alpha)W(\sigma_g) \geq W(\sigma_{\alpha F+(1-\alpha)G}) = W(\alpha\sigma_F + (1-\alpha)\sigma_G)
\end{aligned}$$

$\square$

Lemma 5. *W is translation invariant.*

Proof. Of Lemma 5. By convexity and normalization

$$\begin{aligned}
W(\sigma_F + k) =& W\left(\alpha\frac{\sigma_F}{\alpha} + (1-\alpha)\frac{k}{(1-\alpha)}\right) \\
\leq& \alpha W\left(\frac{\sigma_F}{\alpha}\right) + (1-\alpha)W\left(\frac{k}{1-\alpha}\right) \\
=& \alpha W\left(\frac{\sigma_F}{\alpha}\right) + (1-\alpha)\frac{k}{1-\alpha}
\end{aligned}$$

since it is true for all $\alpha \in (0,1)$, by continuity, it is true as $\alpha \to 1$, then,

$$W(\sigma_F + k) \leq W(\sigma_F) + k$$

Since $u(X)$ is unbounded above and below, for any $\sigma_F \in \hat{\Phi}$ and $k \in \mathbb{R}$,

$$\begin{aligned} W(\sigma_F + k) \leq & W(\sigma_F) + k = W(\sigma_F + k - k) + k \\ \leq & W(\sigma_F + k) - k + k = W(\sigma_F + k) \end{aligned}$$

Then $W(\sigma_F + k) = W(\sigma_F) + k$ for all $\sigma_F \in \hat{\Phi}$ and $k \in \mathbb{R}$. □

To sum up, W is a monotone, normalized, convex and translation invariant functional, by Cerreia-Vioglio et al. (2014, Prop. 2) it is a niveloid. By the properties in section A and Claim 6 of De Oliveira et al. (2013), W can be rewritten as

$$W(\sigma_F) = \max_{\pi \in \Pi(\hat{p})} \int_{\Delta(\Omega)} \max_{f \in F} \left(\int u \circ f \, dp \right) \pi(dp) - c(\pi)$$

where $c : \Pi(\hat{p}) \to (-\infty, \infty]$ is given by

$$c(\pi) = \sup_{F \in \mathscr{A}} \left[\left(\int u \circ f \, dp \right) \pi(dp) - W(\sigma_F) \right]$$

This concludes the proof of part a).

For part b., the fact that a SL representation implies the axioms is trivial. For the converse direction, ITRU clearly implies DITRU and PERU, hence Lemma 1 and Lemma 2 hold. We can define W and V as in the proof of Theorem 1, then it is sufficient to show that W is linear and an application of the Riesz's representation theorem will give the result. Linearity of W follows from ITRU, indeed for all $F, G \in \mathscr{A}$ and all $\alpha \in (0,1)$,

$$\begin{aligned} & \alpha\delta_F + (1-\alpha)\delta_G \sim \delta_{\alpha F + (1-\alpha)G} \iff \\ & \alpha V(F) + (1-\alpha)V(G) = V(\alpha F + (1-\alpha)G) \iff \\ & \alpha W(\sigma_F) + (1-\alpha)W(\sigma_g) = W(\sigma_{\alpha F + (1-\alpha)G}) = W(\alpha\sigma_F + (1-\alpha)\sigma_G) \end{aligned}$$

where the second equality follows from ITRU. By Riesz's representation theorem, there exists a unique $\pi \in \Delta(\Delta(\Omega))$ such that

$$W(\sigma_F) = \int \sigma_F(p)\pi(dp)$$

This concludes the proof of part b).

□

Proof. Of Corollary 1. The fact that CIA implies the axioms is straightforward. To prove the other implication, we follow the proof of Theorem 1. First, we prove that WDI, together with the other axioms, implies the independence axiom. Suppose $\delta_f \sim \delta_g$ and for some $h \in \mathscr{F}$, $\delta_{\frac{1}{2}f+\frac{1}{2}h} \not\sim \delta_{\frac{1}{2}g+\frac{1}{2}h}$. Let assume w.l.o.g. $\delta_{\frac{1}{2}f+\frac{1}{2}h} \succ \delta_{\frac{1}{2}g+\frac{1}{2}h}$, Weak Degenerate Independence implies $\delta_f = \delta_{\frac{1}{2}f+\frac{1}{2}f} \succ \delta_{\frac{1}{2}f+\frac{1}{2}g} \succ \delta_{\frac{1}{2}g+\frac{1}{2}g} = \delta_g$, a contradiction to the initial assumption. Therefore, Lemma 1 holds. The rest of the proof is identical to the one of Theorem 1. □

Before proving Proposition 1, notice that $\partial V(F)$ is equal to the subdifferential of $W(\sigma_G)$ (see De Oliveira et al., 2013).

Proof. Of Proposition 1. (1.) implies (2.). Assume $\pi^* \in \partial W(\sigma_F) \cap \partial W(\sigma_G) \neq \emptyset$, by the definition of subdifferential

$$W(\sigma) \geq W(\sigma_F) + \langle \sigma - \sigma_F, \pi^* \rangle \text{ and } W(\sigma) \geq W(\sigma_G) + \langle \sigma - \sigma_G, \pi^* \rangle \tag{3}$$

for all $\sigma \in C(\Delta(\Omega))$. Then, choosing $\sigma = \sigma_G$ in the first inequality and $\sigma = \sigma_F$ in the second gives:

$$W(\sigma_F) - W(\sigma_G) = \langle \sigma_F - \sigma_G, \pi^* \rangle \tag{4}$$

Now, take $\sigma = \alpha\sigma_F + (1-\alpha)\sigma_G$ for some $\alpha \in (0,1)$ and plug it into the right side of Eq. (3), then

$$\begin{aligned} W(\alpha\sigma_F + (1-\alpha)\sigma_G) \geq & W(\sigma_G) + \alpha\langle \sigma_F - \sigma_G, \pi^* \rangle \\ = & W(\sigma_G) + \alpha\,(W(\sigma_F) - W(\sigma_G)) \\ = & \alpha W(\sigma_F) + (1-\alpha)W(\sigma_G) \end{aligned}$$

where the first equality follows from Eq. (4). The reverse inequality follows from convexity of W.

(2.) implies (1.) Assume 2. and, by contrapositive, $\partial W(\sigma_F) \cap \partial W(\sigma_G) = \emptyset$. Take $\pi \in \partial W(\alpha\sigma_F + (1-\alpha)\sigma_G)$ who is non-empty by Ergin and Sarver (2010, Lemma 2.5), then $\pi \notin \partial W(\sigma_F)$ or $\pi \notin \partial W(\sigma_G)$ by the initial assumption, so $W(\sigma_F) \geq \langle \sigma_F, \pi \rangle - W^*(\pi)$ and $W(\sigma_G) \geq \langle \sigma_G, \pi \rangle - W^*(\pi)$ and one of these inequalities must be strict. Then, for all $\alpha \in (0,1)$:

$$\begin{aligned} \alpha W(\sigma_F) + (1-\alpha)W(\sigma_G) > & \alpha\langle \sigma_F, \pi \rangle + (1-\alpha)\langle \sigma_G, \pi \rangle - W^*(\pi) \\ = & \langle \alpha\sigma_F + (1-\alpha)\sigma_G, \pi \rangle - W^*(\pi) \\ = & W(\alpha\sigma_F + (1-\alpha)\sigma_G) \end{aligned}$$

a contradiction to (2.) and where the last equality follows from $\pi \in \partial W(\alpha\sigma_F + (1-\alpha)\sigma_G)$. □

Proof. Of Theorem 2. The equivalence of 2. and 3. is due to De Oliveira et al. (2013, Prop. 3). We prove the equivalence of 1. and 3. Let start with 3. implies 1., $\partial V_2(F) \cap \partial V_2(G) = \emptyset$ implies, by Prop. 1 and PERU, that $\alpha_F + (1-\alpha)\delta_G \succ_2 \delta_{\alpha F + (1-\alpha)G}$. By 3., this implies $\partial V_1(F) \cap \partial V_1(G) = \emptyset$, hence $\alpha_F + (1-\alpha)\delta_G \succ_1 \delta_{\alpha F + (1-\alpha)G}$. To prove that 1. implies 3. assume that 1. holds and 3. does not hold, then for some $F, G \in \mathscr{A}$, $\partial V_2(F) \cap \partial V_2(G) = \emptyset$ and $\partial V_1(F) \cap \partial V_1(G) \neq \emptyset$. The first condition implies again $\alpha_F + (1-\alpha)\delta_G \succ_2 \delta_{\alpha F + (1-\alpha)G}$, whereas the second condition implies, by Proposition 1, $\alpha_F + (1-\alpha)\delta_G \sim_1 \delta_{\alpha F + (1-\alpha)G}$ a contradiction to the initial hypothesis. □

B.1 Calculation for Example 1

Suppose $F \sim G$, then

$$\begin{aligned} V(\gamma F + (1-\gamma)G) = & \max_{\pi \in \Pi(\hat{p})} \frac{\int \max_{f \in \gamma F + (1-\gamma)G} \left(\int u(f)dp\right) d(\pi(p))}{c(\pi)} \\ = & \max_{\pi \in \Pi(\hat{p})} \left(\gamma \frac{\int \max_{f \in F}(u(f)dp)\pi(dp)}{c(\pi)} + (1-\gamma)\frac{\max_{f \in G}(\int u(f)dp)\pi(dp)}{c(\pi)}\right) \\ \leq & \gamma V(F) + (1-\gamma)V(G) = V(F) \end{aligned}$$

then, V satisfies Aversion to Randomization.

References

Caplin, A. and M. Dean (2014). Revealed preference, rational inattention, and costly information acquisition. Working Paper 19876, National Bureau of Economic Research.

Cerreia-Vioglio, S., F. Maccheroni, M. Marinacci, and L. Montrucchio (2011). Uncertainty averse preferences. *Journal of Economic Theory 146*(4), 1275–1330.

Cerreia-Vioglio, S., F. Maccheroni, M. Marinacci, and A. Rustichini (2014). Niveloids and their extensions: Risk measures on small domains. *Journal of Mathematical Analysis and Applications 413*(1), 343 – 360.

de Oliveira, H. (2014). Axiomatic foundations for entropic costs of attention. *Mimeo.*

De Oliveira, H., T. Denti, M. Mihm, and M. K. Ozbek (2013). Rationally inattentive preferences. *Available at SSRN 2274286.*

Dillenberger, D., J. S. Lleras, P. Sadowski, and N. Takeoka (2014). A theory of subjective learning. *Journal of Economic Theory 153*(0), 287 – 312.

Ellis, A. (2013). Foundations for optimal inattention. *Mimeo.*

Ergin, H. and T. Sarver (2010). The unique minimal dual representation of a convex function. *Journal of Mathematical Analysis and Applications 370*(2), 600–606.

Ergin, H. and T. Sarver (2014). Hidden actions and preferences for timing of resolution of uncertainty. *Theoretical Economics*, Forthcoming.

Kacperczyk, M., S. Van Nieuwerburgh, and L. Veldkamp (2009). Rational attention allocation over the business cycle. NBER Working Papers 15450, National Bureau of Economic Research, Inc.

Kreps, D. M. and E. L. Porteus (1978). Temporal resolution of uncertainty and dynamic choice theory. *Econometrica 46*(1), pp. 185–200.

Luo, Y. (2008). Consumption dynamics under information processing constraints. *Review of Economic Dynamics 11*(2), 366–385.

Mackowiak, B. and M. Wiederholt (2009). Optimal sticky prices under rational inattention. *American Economic Review 99*(3), 769–803.

Masatlioglu, Y., D. Nakajima, and E. Y. Ozbay (2012). Revealed attention. *American Economic Review 102*(5), 2183–2205.

Matějka, F. and A. McKay (2014). Rational inattention to discrete choices: a new foundation for the multinomial logit model. *American Economic Review*, Forthcoming.

Mondria, J. (2010). Portfolio choice, attention allocation, and price comovement. *Journal of Economic Theory 145*(5), 1837–1864.

Sims, C. A. (1998). Stickiness. *Carnegie-Rochester Conference Series on Public Policy 49*(1), 317–356.

Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics 50*(3), 665–690.

Tutino, A. (2008). The rigidity of choice: lifetime savings under information-processing constraints. MPRA Paper 16744, University Library of Munich, Germany.

Van Nieuwerburgh, S. and L. Veldkamp (2009). Information immobility and the home bias puzzle. *Journal of Finance 64*(3), 1187–1215.