KIER DISCUSSION PAPER SERIES

KYOTO INSTITUTE
OF
ECONOMIC RESEARCH

Discussion Paper No.1040

"Costly Subjective Learning"

Youichiro Higashi, Kazuya Hyogo, and Norio Takeoka

August 2020

KYOTO UNIVERSITY

KYOTO, JAPAN

Costly Subjective Learning*

Youichiro Higashi, Kazuya Hyogo, and Norio Takeoka †

August 16, 2020

Abstract

Information acquisition is an important aspect of decision making. Acquiring information is costly, but the cost of information acquisition is not typically observable and hence it is not obvious how it can be measured. Using preference over menus, de Oliveira, Denti, Mihm, and Ozbek [15] provide an axiomatic foundation for the additive costs model of information acquisition. If obtaining signals from experiments is time-consuming, such as in the case of a long-run investment decision, however, costs may be measured as a discount factor or waiting time for acquiring information. We propose a general class of representations which allows for non-additive costs for information acquisition and provide its axiomatic foundation. Furthermore, the discounting costs model is characterized as a special case.

Keywords: costly information acquisition, rational inattention, waiting costs, preference for flexibility, preference over menus.

JEL classification: D11, D81, D91.

*A part of this paper was previously circulated under the title "Information Acquisition with Subjective Waiting Costs". The authors are grateful to Larry Epstein, Jingyi Xue, David Dillenberger, and the audience at RUD 2019 (Paris School of Economics), AMES 2019 (Xiamen University), and seminar participants at Ecole Polytechnique, Hitotsubashi University, Kyoto University, National Taiwan University, Singapore Management University, and University of Paris 1 for their helpful comments. The authors gratefully acknowledge financial support by 2018–2020 Joint Research Program of KIER (Kyoto University), and Open Research Area (ORA) for the Social Sciences 2018. Higashi, Hyogo and Takeoka gratefully acknowledge financial support by Grant-in-Aid for Scientific Research (C).

†Higashi is at the Faculty of Economics, Okayama University, 3-1-1 Tsushima-naka, Kita-ku, Okayama 700-8530, Japan, higashi-y@okayama-u.ac.jp; Hyogo is at the Faculty of Economics, Ryukoku University, 67 Fukakusa Tsukamoto-cho, Fushimi-ku, Kyoto 612-8577, Japan, hyogo@econ.ryukoku.ac.jp; Takeoka is at the Department of Economics, Hitotsubashi University, 2-1 Naka, Kunitachi, Tokyo 186-8601, Japan, norio.takeoka@r.hit-u.ac.jp.

1 Introduction

Information acquisition is an important aspect of decision making. Imagine an economic agent who determines to what extent the agent gets informed and optimally solves a trade-off between benefits obtained from learning an additional piece of information and its associated costs. As in the growing literature of rational inattention (Sims [29]), such a decision making has been recognized increasingly in economics. A feature of this literature is that various implications of the model crucially depends on the specification of costs for acquiring information. However, the cost for information acquisition is not typically observable and hence it is not obvious how it can be measured.

There are two approaches for identifying costs of acquiring information. One is to consider a stochastic choice as a primitive. This approach is taken by Caplin and Dean [4]. Consider a situation where an agent receives a subjective signal, and afterwards, makes a choice from a feasible set (called a menu). Since signals arrive stochastically, observed choices from the menu also becomes stochastic. Hence, by starting with stochastic choice across various menus, an information structure and the associated costs from information acquisition behind the choice may be revealed from the choice data.

The other is to consider preference over menus. This approach is taken by de Oliveira, Denti, Mihm, and Ozbek [15]. Consider a situation where an agent chooses a menu. Presumably, during this choice, he anticipates a costly information acquisition and a subsequent choice from the menu after the arrival of new signals. Therefore, this kind of data are relevant for making inference about costs of information acquisition. The stochastic choice approach considers ex post choices after arriving subjective signals, while the preference-over-menus approach considers ex ante menu choice before arrival of signals. Thus, the two approaches are complementary to each other.

Though the existing literature of information acquisition, both in applications and foundations, typically assumes additive costs, there are some instances where implications of additive information costs are not necessarily reasonable. For example, if obtaining signals from experiments is time-consuming, such as in the case of a long-run investment decision, costs may be measured as a discount factor or waiting time for acquiring information. Chambers, Liu, and Rehbeck [9] take the stochastic choice approach and identify non-additive information costs and in particular, multiplicative costs. The latter is particularly interesting because it can be interpreted as discounting costs. In the present paper, we take the preference-over-menus approach and identify unique (possibly non-additive) costs of information acquisition. Thus, our result is a generalization of de Oliveira, Denti, Mihm, and Ozbek [15] and complementary to Chambers, Liu, and Rehbeck [9].

More precisely, we consider the following model. Let Ω be a finite set of objective states and X be a set of lotteries. A function $f : \Omega \to X$ is called an act. We consider a finite subset F of acts, called a menu, as a choice object. Suppose that the agent has an expected utility function $u : X \to \mathbb{R}$ and an initial prior $\overline{p}$ over Ω. Before making a choice from a menu F, the agent may conduct an additional experiment or engage in information acquisition, which generates signals about states. The agent updates his prior and makes a choice from the menu contingent upon posteriors. Formally, information acquisition is interpreted as a choice of an information structure $\pi \in \Delta(\Delta(\Omega))$ whose prior coincides with $\overline{p}$.

Given each menu F, the value of information of π is defined as[1]

$$b_F^u(\pi) \equiv \int \left(\max_{f \in F} \sum_{\omega} u(f(\omega))p(\omega) \right) \mathrm{d}\pi(p).$$

After choosing the information structure, the agent observes a signal and updates his prior belief to the posterior p. Given the posterior belief p, the agent chooses an act f from a menu F to maximize the expected utility. The value of information is computed as the expectation of these maximum values with respect to the distribution over signals, given by π.

We axiomatize the following representation: A preference $\succsim$ over menus admits a Costly Subjective Learning Representation if there exist an expected utility function $u : X \to \mathbb{R}$, a prior belief $\overline{p}$ over Ω, a function $W(\pi, t)$, interpreted as a net benefit function of information acquisition, such that

$$U(F) = \max_{\pi \in \Pi(\overline{p})} W(\pi, b_F^u(\pi)) \tag{1}$$

represents $\succsim$, where $\Pi(\overline{p})$ is the set of information structures consistent with the prior. This representation is given as an indirect utility function of the maximization where the agent optimally chooses an information structure by considering benefits and costs of acquiring information. We impose reasonable properties on W, which justify our interpretation of W being a net benefit of information acquisition.

In the Costly Subjective Learning Representation, costs for information acquisition are implicitly incorporated into W. A special case is the rationally inattentive representation, considered in de Oliveira, Denti, Mihm, and Ozbek [15]:

$$U(F) = \max_{\pi \in \Pi(\overline{p})} (b_F^u(\pi) - c(\pi)), \tag{2}$$

where $c(\pi)$ is a cost function to choose π. It is easy to see that (2) is obtained from (1) with specifying $W(\pi, t) = t - c(\pi)$. In terms of axiomatization, this class of representations is characterized by adding one more axiom to (1).

By adding yet another axiom to (1), we also characterize an alternative specification of the Costly Subjective Learning Representation. A preference $\succsim$ over menus admits an optimal waiting representation if there exist a discounting function $\beta(\pi) \in [0, 1]$, and a premium function $\gamma(\pi) \in [1, \infty]$ such that

$$U(F) = \max_{\pi \in \Pi(\overline{p})} [\beta(\pi)(b_F^u(\pi))^+ - \gamma(\pi)(b_F^u(\pi))^-]$$

represents $\succsim$, where $(t)^+ = \max\{0, t\}$, $(t)^- = \max\{0, -t\}$ for $t \in \mathbb{R}$, and $\infty \times 0 = 0$ with convention. If $b_F^u(\pi)$ is positive, it is discounted by $\beta(\pi)$, while if $b_F^u(\pi)$ is negative, the

[1] We borrow the same notation from de Oliveira, Denti, Mihm, and Ozbek [15].

loss is amplified by the premium rate $\gamma(\pi)$. In this representation, the agent behaves as if he optimally chooses an information structure π by taking into account trade-offs with waiting time, captured by discounting $\beta(\pi)$ or premium rate $\gamma(\pi)$.

To prove the main theorem, we borrow techniques from the literature of choice under ambiguity. Indeed, the additive cost model satisfies the property, called translation invariance, and has a parallel relationship with the variational representation of Maccheroni, Marinacci, and Rustichini [27]. On the other hand, the optimal waiting representation satisfies homotheticity and has a parallel relationship with the confidence representation of Chateauneuf and Faro [10]. The Costly Subjective Learning Representation nests both classes, and is a counterpart of the uncertain averse representation of Cerreia-Vioglio, Maccheroni, Marinacci, and Montrucchio [8].

1.1 Implications of additive and discounting costs

For simplicity, assume that the objective state space is given by $\Omega = \{\omega_1, \omega_2\}$. In this illustration, an act is a function defined on Ω which pays a positive payoff according to a realization of states. Hence, an act is identified with (x_1, x_2), where $x_i \in \mathbb{R}$ is a payoff when ω_i is realized, and is interpreted as an investment. Payoffs are interpreted as either monetary prizes with assuming the agent's being risk neutral or utils.

Imagine an agent who faces with a decision about investment opportunities, which are regarded as menus of acts. The agent has preference over menus of acts.

Suppose that

$$\{(100, 0)\} \sim \{(0, 100)\} \sim \{(50, 50)\}.$$

The first indifference ranking implies that the agent's prior over states is given by $(\frac{1}{2}, \frac{1}{2})$. The second indifference ranking suggests that his willingness to pay to each act is 50. Suppose that a menu $\{(100, 0), (0, 100)\}$ is also available. If this menu is chosen, the agent can postpone his investment decision in the future. If a choice is made from $\{(100, 0), (0, 100)\}$ based on his prior, the willingness to pay for such a choice is still 50. But, presumably, facing with the menu $\{(100, 0), (0, 100)\}$, the agent optimally solves a problem of costly information acquisition, in which case the agent can make a decision contingent upon arrivals of new information. If the agent anticipates such an information acquisition, he may exhibit preference for flexibility such as

$$\{(100, 0), (0, 100)\} \sim \{(60, 60)\} \succ \{(100, 0)\} \sim \{(0, 100)\}.$$

If evaluated at an optimal information structure, $\{(100, 0), (0, 100)\}$ is worth 60. Therefore, the marginal (net) benefit of acquiring this information structure is given as 10.

Let us consider an implication of the additive cost model given as in (2). Since the additive cost model satisfies the property, called translation invariance, for all positive m,

$$\{(100 + m, m)\} \sim \{(m, 100 + m)\} \sim \{(50 + m, 50 + m)\},$$
$$\text{and } \{(100 + m, m), (m, 100 + m)\} \sim \{(60 + m, 60 + m)\}.$$

These rankings suggest that the marginal (net) benefit of acquiring a new information structure is still given as 10, which implies that an optimal level of information acquisition is invariant between $\{(100,0),(0,100)\}$ and $\{(100+m,m),(m,100+m)\}$ for all $m>0$.

However, a level of common payoff m may affect an incentive for costly information acquisition. If the agent anticipates that information search is a time-consuming task, costs for information acquisition come from time delay of decisions or discounting associated with the waiting time. On the one hand, if m is sufficiently large, the significance of the state-dependent payoff of 100 relative to the constant payoff m seems to be diminished. On the other hand, if the decision is delayed by information acquisition, the constant payoff m is also delayed and this cost from waiting becomes more significant when m is larger. Thus, the agent may become less willing to acquire a new information structure and quit information acquisition sooner. Consequently, he may exhibit

$$\{(60+m,60+m)\} \succ \{(100+m,m),(m,100+m)\} \succ \{(50+m,50+m)\}$$

for large $m>0$. Note that $\{(60+m,60+m)\}$ is indifferent to an alternative where the agent is supposed to choose between $(100,0)$ and $(0,100)$ and is given m immediately. On the other hand, when the agent faces with $\{(100+m,m),(m,100+m)\}$, he solves an optimal waiting time by considering both information acquisition for state-dependent payoff of 100 and costs from delaying a constant payoff m. He will end up with a sub-optimal level of information acquisition (compared with the case of $\{(100,0),(0,100)\}$) and some delay to obtain m. Thus, $\{(100+m,m),(m,100+m)\}$ is less preferred to $\{(60+m,60+m)\}$, as stated above.

In the case of an optimal waiting representation as given in (1),

$$\begin{aligned}
& U(\{(100+m,m),(m,100+m)\}) \\
= & \max_{\pi} \beta(\pi) b_{\{(100+m,m),(m,100+m)\}}(\pi) = \max_{\pi} \beta(\pi)(b_{\{(100,0),(0,100)\}}(\pi)+m) \\
= & \beta(\pi^*) b_{\{(100,0),(0,100)\}}(\pi^*) + \beta(\pi^*) m \leq \beta(\pi^*) b_{\{(100,0),(0,100)\}}(\pi^*) + m \\
\leq & \max_{\pi} \beta(\pi) b_{\{(100,0),(0,100)\}}(\pi) + m = U(\{(100,0),(0,100)\}) + m \\
= & U(\{(60,60)\}) + m = U(\{(60+m,60+m)\}).
\end{aligned}$$

Moreover,

$$\begin{aligned}
& U(\{(100+m,m),(m,100+m)\}) \\
= & \max_{\pi} \beta(\pi) b_{\{(100+m,m),(m,100+m)\}}(\pi) = \max_{\pi} \beta(\pi)(b_{\{(100,0),(0,100)\}}(\pi)+m) \\
\geq & \beta(\delta_{\overline{p}})(b_{\{(100,0),(0,100)\}}(\delta_{\overline{p}})+m) = b_{\{(100,0),(0,100)\}}(\delta_{\overline{p}}) + m \\
= & U(\{(50,50)\}) + m = U(\{(50+m,50+m)\})
\end{aligned}$$

because $\beta(\delta_{\overline{p}})=1$, that is, there is no discounting if the choice is made according to the prior information. Therefore, the optimal waiting representation is consistent with the above intuition.

1.2 Related literature

By investigating preferences over pairs consisting of an action and a menu of acts, Hyogo [22] characterizes general models and additive models of costly information acquisition. Though information contents are subjective, he assumes that choices of informations are observable.

In the literature of preference over menus, subjective optimization is introduced by Ergin and Sarver [21] in the context of contemplation costs. They generalize the additive representation of Dekel, Lipman, and Rustichini [12], and characterize the additive cost function for contemplating subjective states, which is technically regarded as a counterpart of the variational preference of Maccheroni, Marinacci, and Rustichini [27].

Dillenberger, Lleras, Sadowski, and Takeoka [18] extend the framework of Dekel, Lipman, and Rustichini [12] by considering preference over menus of acts as in the present paper. They derive a subjective information structure from preference and call their framework the subjective learning. In their framework, the agent uses a single information structure for all menus. To accommodate the menu-dependent aspect of information acquisition, de Oliveira, Denti, Mihm, and Ozbek [15] generalize the subjective learning model and characterize a subjective optimization under additive costs for information acquisition. As a special case of the additive costs, de Oliveira [14] axiomatizes a specific cost function, called the relative entropy, which is commonly used in the literature of rational inattention.

Dillenberger, Krishna, and Sadowski [16, 17] consider repeated decisions of information acquisition in an infinite horizon framework of menu choice. Their main focus is to model information acquisition from a constrained set of information structures. This specific setting only admits a constraint without cost functions, and is called a constrained information model. The agent in this model faces a trade-off between acquiring information and its timing because of discount factors. But, unlike our optimal waiting representation, a discount factor itself is independent of information acquisition.

An alternative approach to identify costs for information acquisition is to consider a stochastic choice from menus of acts. Caplin and Dean [4] identify additive costs for information acquisition from a state-dependent stochastic choice. Caplin, Dean and Leahy [5] and Denti [13] also take a state-dependent stochastic choice as primitives and characterize a specific class of additive costs, called posterior separable costs, which includes the expected relative entropy of posterior and prior beliefs used in the literature of rational inattention. Chambers, Liu, and Rehbeck [9] take the same primitives of Caplin and Dean [4] and identify non-additive information costs and in particular, multiplicative costs. Thus, their result is complementary to our result established in the model of preference over menus.

Lin [25] provides a parsimonious model by only assuming state-independent stochastic choice, which is built on the framework of Lu [26], and characterizes additive costs. Duraj and Lin [19] also take the parsimonious framework and characterize discounting costs.

Ellis [20] considers a state-dependent deterministic choice function from menus of acts and derives a cost function for partitions, which is interpreted as costly partitional learning. Aoyama [1] extends Ellis [20] by incorporating decision time as a part of primitives and derives a cost function for filtrations.

Waiting to invest has been studied by several papers. Kendall [24] considers a similar problem of optimal information structure (rush or wait), but the waiting cost in his model comes from the situation where some piece of information becomes public later. There is no discounting in his model.

Another similar model is an optimal stopping problem. A value of a perpetual American option should be determined by a solution to the optimal stopping problem. In the optimal stopping problem, an information structure, given by a underlining stochastic process $(V_t)_{t=1}^{\infty}$, is given, and an optimal stopping time $\tau = (\tau_t)_{t=0}^{\infty}$, where τ_t is V_t-adaptive, is determined so as to maximize

$$\sup_{\tau} \mathbb{E}\left[e^{-r(\tau)}(V_\tau - I)\right]. \tag{3}$$

On the other hand, in our model, information structures are variable. The decision maker chooses an optimal information structure and the associated waiting time for each decision problem. Another difference is the timing of decision about waiting time. In the optimal stopping time, the agent makes a decision about investment according to realizations of signals. In our model, the decision is made ex ante, that is, the agent makes a commitment to the waiting time no matter what realizations of signals are observed until then. Indeed, if the agent makes a commitment to the waiting time ex ante, he chooses $t \in \{0, 1, 2, \cdots\}$ rather than a stopping time $\tau = (\tau_t)_{t=0}^{\infty}$. In this case, (3) is rewritten as

$$\sup_{t} e^{-r(t)}\mathbb{E}\left[V_t - I\right],$$

which is a special case of our waiting time representation.

In the above model, the fixed cost I for investment may matter for the optimal waiting time. More precisely, if the cost I for investment increases, the agent tends to wait longer. On the other hand, in the additive cost model, I is independent of the optimal waiting time. We will make similar comparisons between the two models in Section 4.4.

2 Costly subjective learning representations

2.1 Primitives

We consider the following as primitives of the model. These primitives are exactly the same as in de Oliveira, Denti, Mihm, and Ozbek [15].

- $\Omega = \{\omega_1, ..., \omega_n\}$: the (finite) objective state space
- X: outcomes, consisting of simple lotteries on a set of deterministic prizes
- $f : \Omega \to X$: an (Anscombe-Aumann) act
- $\mathcal{F}$: the set of all acts
- $F \subset \mathcal{F}$: a finite set of acts, called a menu

- $\mathbb{F}$: the set of all menus
- Preference $\succsim$ over $\mathbb{F}$

2.2 Functional form

Let $\overline{p} \in \Delta(\Omega)$ be the agent's prior belief. A probability distribution $\pi \in \Delta(\Delta(\Omega))$ is interpreted as an information structure or a signal structure about Ω. For each π, the initial prior $p^{\pi} \in \Delta(\Omega)$ associated with π is defined as

$$p^{\pi}(\omega) = \int_{\Delta(\Omega)} p(\omega)\, \mathrm{d}\pi(p)$$

for each ω. We impose a restriction on the relationship between the prior belief and subjectively possible information structures. We say that π satisfies a martingale property or a Bayesian plausibility constraint (Kamenica and Gentzkow [23]) if

$$p^{\pi} = \overline{p}. \tag{4}$$

That is, the initial prior associated with π exactly coincides with the agent's prior belief $\overline{p}$. Define

$$\Pi(\overline{p}) = \{\pi \in \Delta(\Delta(\Omega)) \,|\, p^{\pi} = \overline{p}\},$$

which is weak* closed and convex.

Given $u : X \to \mathbb{R}$ and a menu F, an information value of $\pi \in \Pi(\overline{p})$ is defined as

$$b_F^u(\pi) \equiv \int_{\Delta(\Omega)} \max_{f \in F} \left(\sum_{\omega \in \Omega} u(f(\omega))p(\omega) \right) \mathrm{d}\pi(p).$$

In particular, for any singleton menu $F = \{f\}$ and $\pi \in \Pi(\overline{p})$, we have

$$b_F^u(\pi) = \sum_{\Omega} u(f(\omega))\overline{p}(\omega),$$

that is, the information value exactly coincides with the expected utility of f under the prior if the agents makes a commitment.

To capture benefits from information acquisition, we introduce the Blackwell order, which gives a partial order on $\Delta(\Delta(\Omega))$ in terms of informativeness of signals.

Definition 1 *A signal $\pi \in \Delta(\Delta(\Omega))$ is Blackwell more informative than a signal $\rho \in \Delta(\Delta(\Omega))$, denoted $\pi \trianglerighteq \rho$, if*

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

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

Since $\max_{f\in F}\left(\sum u(f(\omega))p(\omega)\right)$ is convex and continuous in p, we have $b^u_F(\pi) \geq b^u_F(\rho)$ whenever π is Blackwell more informative than ρ.

The value $b^u_F(\pi)$ is the gross value of information, which dose not involve costs for choosing π. For optimal information acquisition, the net value of information is more relevant. We consider a function $W : \Pi(\overline{p}) \times \mathbb{R} \to \mathbb{R} \cup \{-\infty\}$ which captures the net value of information where costs for acquiring information is taken into account. We say that W is linearly continuous if the map

$$\varphi \mapsto \sup_{\pi\in\Pi(\overline{p})} W\left(\pi, \int \varphi \, \mathrm{d}\pi\right)$$

from the set of continuous functions on $\Delta(\Omega)$, denoted by $C(\Delta(\Omega))$, into $[-\infty, \infty]$ is extended-valued continuous. Let δ_a denote the Dirac measure at a.

Definition 2 *We say that $W : \Pi(\overline{p}) \times \mathbb{R} \to \mathbb{R} \cup \{-\infty\}$ is a net benefit function if*

(i) W is quasi-concave, upper semi-continuous, and linearly continuous,

(ii) for all π, $W(\pi, t)$ is non-decreasing in t,

(iii) $W(\delta_{\overline{p}}, t) = t$ for the initial prior $\overline{p}$,

(iv) for all t and $\pi, \rho \in \Pi(\overline{p})$, $\pi \trianglerighteq \rho \implies W(\pi, t) \leq W(\rho, t)$.

The function $W(\pi, t)$ captures the net benefit when an information structure π is chosen and t is the gross value of information. Part (i) is a technical condition to ensure a well-defined optimization problem of information acquisition. Part (ii) states that for each fixed π, the net benefit increases accordingly when the gross value of information increases. Part (iii) states that the gross and net values coincide if the prior information is used. In other words, there is no cost if there is no additional information acquisition. Part (iv) states that a more informative signal structure is more costly. In fact, for each fixed level of t, its net benefit is lower under a more informative signal structure. Note also that any $\pi \in \Pi(\overline{p})$ is a mean-preserving spread of $\delta_{\overline{p}}$, and hence $\pi \trianglerighteq \delta_{\overline{p}}$. From parts (iii) and (iv), $W(\pi, t) \leq W(\delta_{\overline{p}}, t) = t$, which means that the net benefit is always lower than the gross value of information t. Hence, the cost of choosing π is implicitly embodied into W.

Definition 3 *A Costly Subjective Learning Representation is a tuple $(u, \overline{p}, W)$, where $u : X \to \mathbb{R}$ is an unbounded expected utility function with $u(X) = \mathbb{R}$, $\overline{p}$ is the initial prior, and W is a net benefit function such that $\succsim$ is represented by*

$$U(F) = \max_{\pi\in\Pi(\overline{p})} W(\pi, b^u_F(\pi)).$$

2.3 Special cases

- Fixed information model (Dillenberger, Lleras, Sadowski, and Takeoka [18]: There exists $(u, \overline{p}, \pi^*)$ such that $p^{\pi^*} = \overline{p}$ and

  $$U(F) = b_F^u(\pi^*).$$

  In this model, the agent has a fixed information structure, which is applied for all menus. The fixed information model is a special case of the Costly Subjective Learning Representation when $W(\pi, t) = t$ for all t and π with $\pi^* \trianglerighteq \pi$ and $W(\pi, t) = -\infty$ otherwise.

- Constrained information model: There exists $(u, \overline{p}, \Pi)$ such that $\Pi \subset \Pi(\overline{p})$ is weak* closed and convex, and

  $$U(F) = \max_{\pi \in \Pi} b_F^u(\pi).$$

  The constrained information model is a special case of Costly Subjective Learning Representation when for all $\pi' \in \Pi$ and $\pi \in \Pi(\overline{p})$ with $\pi' \trianglerighteq \pi$, $W(\pi, t) = t$, and otherwise $W(\pi, t) = -\infty$.

- Additive cost model (de Oliveira, Denti, Mihm, and Ozbek [15]): There exist $(u, \overline{p}, c)$ such that $c : \Pi(\overline{p}) \to [0, \infty]$ is a cost function and

  $$U(F) = \max_{\pi \in \Pi(\overline{p})} \{b_F^u(\pi) - c(\pi)\}.$$

  This model is a special case of Costly Subjective Learning Representation when $W(\pi, t) = t - c(\pi)$ for all $\pi \in \Pi(\overline{p})$.

- Multiplicative cost (discounting cost) model: There exists $(u, \overline{p}, \beta, \gamma)$ such that $\beta : \Pi(\overline{p}) \to [0, 1]$ is a discount function, $\gamma : \Pi(\overline{p}) \to [1, \infty]$ is a premium function, and

  $$U(F) = \max_{\pi \in \Pi(\overline{p})} [\beta(\pi)(b_F^u(\pi))^+ - \gamma(\pi)(b_F^u(\pi))^-],$$

  where $(t)^+ = \max\{0, t\}$ and $(t)^- = \max\{0, -t\}$ for all $t \in \mathbb{R}$. If $b_F^u(\pi) > 0$, then a discount factor $\beta(\pi)$ is applied for computing its net value. If $b_F^u(\pi) < 0$, then a premium factor $\gamma(\pi)$ is applied for computing its net value (or net loss).

2.4 Behavioral foundation

2.4.1 Basic Axioms

We provide a behavioral foundation of the Costly Subjective Learning Representation. We start with the basic axioms that are consistent with any type of costly information acquisition.

Axiom 1 (Order) $\succsim$ *satisfies completeness and transitivity.*

For all F, G and $\alpha \in [0,1]$, define a mixture of F and G by

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

where $\alpha f + (1-\alpha)g \in \mathcal{F}$ is defined by the state-wise mixture between f and g.

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

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

Axiom 3 (Preference for Flexibility) *For all menus F and G, if $G \subset F$, then $F \succsim G$.*

This axiom states that a bigger menu is always weakly preferred.

Axiom 4 (Dominance) *For all menus F and acts g, if there exists $f \in F$ with $f(\omega) \succsim g(\omega)$ for all $\omega \in \Omega$, then $F \sim F \cup \{g\}$.*

Since $F \subset F \cup \{g\}$, the latter menu is weakly preferred by preference for flexibility. If $f(\omega) \succsim g(\omega)$ for all $\omega \in \Omega$, however, for all states, f gives a preferred lottery than g does. In this sense, g is dominated by f. No matter what belief the agent has on states, g should not be chosen over f. Thus, adding g to F does not provide a strictly higher value of flexibility than F.

Axiom 5 (Two-Sided Unboundedness) *There are outcomes $x, y \in X$ with $\{x\} \succ \{y\}$ such that for all $\alpha \in (0,1)$, there are $z, z' \in X$ satisfying*

$$\{\alpha z' + (1-\alpha)y\} \succ \{x\} \succ \{y\} \succ \{\alpha z + (1-\alpha)x\}.$$

This axiom implies the unbounded range of a utility function over outcomes X.[2]

2.4.2 Substantive axioms for information acquisition

A key observation is that an independence-type axiom is crucial to determine the structure of cost functions associated with information acquisition. As a benchmark, we start with a full independence condition.

Axiom 6 (Independence) *For all menus F, G, H, and $\alpha \in (0,1)$*

$$F \succsim G \iff \alpha F + (1-\alpha)H \succsim \alpha G + (1-\alpha)H.$$

[2] de Oliveira, Denti, Mihm, and Ozbek [15] assume one-sided unboundedness: There are outcomes $x, y \in X$ with $\{x\} \succ \{y\}$ such that for all $\alpha \in (0,1)$, there is $z \in X$ satisfying either $\{\alpha z + (1-\alpha)y\} \succ \{x\}$ or $\{y\} \succ \{\alpha z + (1-\alpha)x\}$. The role of our Axiom 5 is explained in the proof sketch of the theorem. See Section 2.5.

Dillenberger, Lleras, Sadowski, and Takeoka [18] and de Oliveira, Denti, Mihm, and Ozbek [15] show that preference $\succsim$ satisfies the basic axioms and Independence if and only if it admits a fixed information representation, that is, there exists (u, π^*) such that

$$U(F) = b_F^u(\pi^*).$$

In this representation, the agent uses a fixed information structure for all menus. The model does not allow for costly information acquisition. This theorem suggests that we have to relax Independence to accommodate costly information acquisition.

The first weakening is to impose Independence only on singleton menus.

Axiom 7 (Singleton Independence) *For all acts f, g, h, and $\alpha \in (0,1)$*

$$\{f\} \succsim \{g\} \iff \alpha\{f\} + (1-\alpha)\{h\} \succsim \alpha\{g\} + (1-\alpha)\{h\}.$$

If the agent makes a commitment to a singleton menu $\{f\}$, there is no role for information acquisition after menu choice. Thus, the commitment rankings reflect the agent's prior belief over states. Singleton Independence implies that the agent follows the subjective expected utility to evaluate acts with commitment according to his prior belief.

Formally, the next axiom requires quasi-convexity of preference.

Axiom 8 (Aversion to Contingent Planning) *For all menus F and G and $\alpha \in (0,1)$,*

$$F \sim G \implies F \succsim \alpha F + (1-\alpha)G.$$

Note that $\alpha F + (1-\alpha)G$ is the menu of contingent plans of the form $\alpha f + (1-\alpha)g$, where $f \in F$ and $g \in G$. If the agent has $\alpha F + (1-\alpha)G$, the randomization α is realized after the agent makes a choice from $\alpha F + (1-\alpha)G$. Thus, information acquisition can not be completely tailored for F and G. The axiom states that the agent avoids contingent planning.

Now we are ready for providing a representation theorem.

Theorem 1 *Preference $\succsim$ satisfies the basic axioms, Singleton Independence, and Aversion to Contingent Planning if and only if it admits a Costly Subjective Learning Representation $(u, \overline{p}, W)$. Moreover, the net benefit function W is obtained as*

$$W(\pi, t) = \inf_{\{F \mid b_F^u(\pi) \geq t\}} u(x_F), \tag{5}$$

where $x_F \in X$ is a lottery equivalent of F satisfying $F \sim \{x_F\}$.

The expression of W given as in (5) provides an explicit formula for eliciting the net benefit function. The expected utility function u is elicited in a standard way. Then, the gross value of information $b_F^u(\pi)$ is computed according to its definition. If a lottery equivalent of each F is elicited from the agent's preference, the net benefit function, under which costs for information acquisition is implicitly involved, can be computed according to (5).

The next theorem shows the uniqueness property of the Costly Subjective Learning Representation.

Theorem 2 *If there exist two Costly Subjective Learning Representations of $\succsim$, denoted by $(u_i, \overline{p}_i, W_i)$ for $i = 1, 2$, then there exist $\alpha > 0$ and $\beta \in \mathbb{R}$ such that $u_2 = \alpha u_1 + \beta$, $\overline{p}_1 = \overline{p}_2 = \overline{p}$, and*

$$W_2(\pi, t) = \alpha W_1\left(\pi, \frac{t - \beta}{\alpha}\right) + \beta$$

for all $(t, \pi) \in \Pi(\overline{p}) \times \mathbb{R}$.

Proof. By the uniqueness result of Anscombe and Aumann [2], there exist $\alpha > 0$ and $\beta \in \mathbb{R}$ such that $u_2 = \alpha u + \beta$. Moreover, $\overline{p}_1 = \overline{p}_2 = \overline{p}$. For any $(\pi, t) \in \Pi(\overline{p}) \times \mathbb{R}$,

$$\begin{aligned} W_2(\pi, t) &= \inf_{\{F \,|\, b_F^{u_2}(\pi) \geq t\}} u_2(x_F) = \inf_{\{F \,|\, \alpha b_F^{u_1}(\pi) + \beta \geq t\}} \alpha u_1(x_F) + \beta \\ &= \alpha \inf_{\{F \,|\, b_F^{u_1}(\pi) \geq \frac{t-\beta}{\alpha}\}} u_1(x_F) + \beta = \alpha W_1\left(\pi, \frac{t - \beta}{\alpha}\right) + \beta. \end{aligned}$$

■

2.5 Proof sketch of Theorem 1

The following is a proof sketch of the sufficiency. We first establish a representation over singleton menus. Singleton Independence together with the basic axioms ensures that $\succsim$ on the singleton menus satisfies all the axioms of Anscombe-Aumann model. There exists an expected utility representation $u : X \to \mathbb{R}$ with unbounded range and a prior $\overline{p} \in \Delta(\Omega)$ such that $\succsim$ over $\mathcal{F}$ is represented by

$$U(f) = \sum_{\Omega} u(f(\omega))\overline{p}(\omega).$$

The property of unbounded range is implied by the Two-Sided Unboundedness axiom.

The axioms ensure the existence of lottery equivalent for all menus. That is, for all F, there exists $x_F \in X$ such that $\{x_F\} \sim F$. The representation $U : \mathcal{F} \to \mathbb{R}$ is extended to $\mathbb{F}$ by $U(F) = U(x_F)$.

For any $F \in \mathbb{F}$ and any $p \in \Delta(\Omega)$, let

$$\varphi_F(p) = \max_{f \in F} \sum_{\Omega} u(f(\omega))p(\omega).$$

This function $\varphi_F : \Delta(\Omega) \to \mathbb{R}$ is interpreted as the support function of F. Let $\Phi_{\mathbb{F}} = \{\varphi_F \,|\, F \in \mathbb{F}\} \subset C(\Delta(\Omega))$ be the set of all support functions. An important property of the support function is that it identifies the menu up to indifference, that is,

$$\varphi_F = \varphi_G \implies F \sim G.$$

Given this identification, we can define the functional $V : \Phi_{\mathbb{F}} \to \mathbb{R}$ by $V(\varphi_F) = U(F)$. Since $\succsim$ satisfies Mixture Continuity and Aversion to Contingent Planning, we can show that V is monotone, normalized, quasi-convex, and continuous by following the techniques of Cerreia-Vioglio, Maccheroni, Marinacci, and Montrucchio [8]. The functional V is extended to the set of all continuous functions $C(\Delta(\Omega))$ by

$$V(\varphi) = \inf\{ V(\varphi_F) \,|\, \varphi_F \in \Phi_{\mathbb{F}}, \varphi_F \geq \varphi\}$$

for all $\varphi \in C(\Delta(\Omega))$ with preserving the above properties. Let $ca_+(\Delta(\Omega))$ be the set of non-negative measures on $\Delta(\Omega)$. For all $\pi \in ca_+(\Delta(\Omega))$ and $t \in \mathbb{R}$, define

$$\begin{aligned} B(\pi, t) &= \{\varphi \in C(\Delta(\Omega)) \,|\, \langle \varphi, \pi \rangle \geq t\}, \text{ and} \\ W(\pi, t) &= \inf_{\varphi \in B(\pi,t)} V(\varphi), \end{aligned}$$

where $\langle \varphi, \pi \rangle = \int \varphi(p)\, \mathrm{d}\pi(p)$. Since all constant functions belong to $C(\Delta(\Omega))$, $B(\pi, t) \neq \emptyset$ for all π and t. Thus, $W(\pi, t) < \infty$ for all (π, t), but it is possible that $W(\pi, t) = -\infty$ for some (π, t). Moreover, since $B(\pi, t)$ is homogeneous of degree zero, so is $W(\pi, t)$.

Now, by definition of B, $\varphi \in B(\pi, \langle \varphi, \pi \rangle)$ for all φ. By definition of W, $V(\varphi) \geq W(\pi, \langle \varphi, \pi \rangle)$ for all φ, which in turn implies

$$V(\varphi) \geq \sup_{\pi \in ca_+(\Delta(\Omega))} W(\pi, \langle \varphi, \pi \rangle),$$

for all φ. A critical step is to show that there exists $\pi \in ca_+(\Delta(\Omega))$ which exactly achieves this supremum, that is, we have

$$V(\varphi) = \max_{\pi \in ca_+(\Delta(\Omega))} W(\pi, \langle \varphi, \pi \rangle).$$

Since W is homogeneous of degree zero, the above maximum is achieved on $\Delta(\Delta(\Omega))$. Moreover, the domain of the maximization problem can be restricted to

$$\Pi = \{\pi \in \Delta(\Delta(\Omega)) \,|\, W(\pi, t) > -\infty \text{ for some } t\}.$$

Since $\langle \varphi_F, \pi \rangle = b_F^u(\pi)$ for all menus F, $\succsim$ is represented by

$$U(F) = V(\varphi_F) = \max_{\pi \in \Pi} W(\pi, b_F^u(\pi)). \tag{6}$$

The remaining step is to show the desired properties of W. A key property is the Bayesian plausibility condition, that is, $\Pi \subset \Pi(\overline{p})$, which has no counterpart in the literature of ambiguity. If it is the case, the maximization of (6) is taken on $\Pi(\overline{p})$ with additionally requiring $W(\pi, \cdot) = -\infty$ for $\pi \in \Pi(\overline{p}) \setminus \Pi$. The Bayesian plausibility is established by unboundedness of u.[3] In fact, if there exists $\pi^* \in \Pi$ such that $p^{\pi^*} \neq \overline{p}$, we can find two

[3] In de Oliveira, Denti, Mihm, and Ozbek [15], the Bayesian plausibility is shown as follows: Take any $\pi \in \Pi$. For any natural number n and $\omega \in \Omega$, take an act f such that $u(f(\omega)) = n$ and $u(f(\omega')) = 0$ for all $\omega' \neq \omega$. In case of the additive cost representation,

$$\langle \varphi_{\{f\}}, \pi \rangle - c(\pi) = np^\pi(\omega) - c(\pi) \leq V(\varphi_{\{f\}}) = n\overline{p}(\omega),$$

which implies $p^\pi(\omega) - \frac{c(\pi)}{n} \leq \overline{p}(\omega)$. As long as $c(\pi) < \infty$, $p^\pi(\omega) \leq \overline{p}(\omega)$ as $n \to \infty$. Since this is true for all ω, we have $p^\pi = \overline{p}$. However, this argument goes through only for the additive cost model. We need a novel manner for the general model, which is shown in the present paper.

states ω and ω' such that $p^{\pi^*}_{\omega} > \overline{p}_{\omega}$ and $p^{\pi^*}_{\omega'} < \overline{p}_{\omega'}$. Then, as shown below, we can find two acts f, $\widetilde{f}$ such that $\{f\} \sim \{\widetilde{f}\}$ but $U(\{f\}) < U(\{\widetilde{f}\})$, which is a contradiction.

By definition of Π, there exists some t^* such that $W(\pi^*, t^*) > -\infty$. Take any $a < W(\pi^*, t^*)$. Let f be a constant act satisfying $u(f(\omega)) = a$ for all ω. Let $\mathbf{a}$ denote the vector in $\mathbb{R}^{\Omega}$ which takes a value of a for all coordinates. Take an alternative vector $\widetilde{\mathbf{a}} \in \mathbb{R}^{\Omega}$ which differs from $\mathbf{a}$ only in the above two coordinates ω and ω', that is,

$$\widetilde{\mathbf{a}} = (a, \cdots, a, \widetilde{a}_{\omega}, a, \cdots, a, \widetilde{a}_{\omega'}, a, \cdots, a) \in \mathbb{R}^{\Omega}.$$

Since $u(X) = \mathbb{R}$, we can find some $\widetilde{f} \in \mathcal{F}$ satisfying $u(\widetilde{f}) = \widetilde{\mathbf{a}}$. Moreover, assume $\widetilde{\mathbf{a}} \cdot \overline{p} = a$. Thus, on the one hand, we have $U(\{f\}) = \mathbf{a} \cdot \overline{p} = a = \widetilde{\mathbf{a}} \cdot \overline{p} = U(\{\widetilde{f}\})$.

On the other hand, since $p^{\pi^*}_{\omega} > \overline{p}_{\omega}$ and $p^{\pi^*}_{\omega'} < \overline{p}_{\omega'}$, $\widetilde{\mathbf{a}} \cdot p^{\pi^*}$ varies across all the real numbers. By choosing $\widetilde{a}_{\omega}$ appropriately, we can set $\widetilde{\mathbf{a}} \cdot p^{\pi^*} = t^*$. By assumption,

$$U(\{f\}) = a < W(\pi^*, t^*) = W(\pi^*, \widetilde{\mathbf{a}} \cdot p^{\pi^*}) \leq \max_{\pi \in \Pi} W(\pi, \widetilde{\mathbf{a}} \cdot p^{\pi}) = U(\{\widetilde{f}\}),$$

which contradicts to the representation.

2.6 Interpersonal comparisons

Consider two agents $i = 1, 2$ having preferences $\succsim_i$ on $\mathbb{F}$. The following condition is a behavioral comparison in terms of attitude toward flexibility. The same condition is considered in Ergin and Sarver [21], Dillenberger, Lleras, Sadowski, and Takeoka [18], and de Oliveira, Denti, Mihm, and Ozbek [15].

Definition 4 *$\succsim_1$ is more averse to commitment than $\succsim_2$ if for all $F \in \mathbb{F}$ and $f \in \mathcal{F}$,*

$$F \succsim_2 \{f\} \implies F \succsim_1 \{f\}.$$

We have the following characterization.

Theorem 3 *Given two preferences $\succsim_i$, $i = 1, 2$ with Costly Subjective Learning Representations $(u_i, \overline{p}_i, W_i)$ for $i = 1, 2$, the following conditions are equivalent:*

(a) $\succsim_1$ is more averse to commitment than $\succsim_2$;

(b) there exist $\alpha > 0$ and $\beta \in \mathbb{R}$ such that $u_2 = \alpha u_1 + \beta$, $\overline{p}_1 = \overline{p}_2 = \overline{p}$, and $W_1(\pi, t) \geq W_2(\pi, t)$ for all $(\pi, t) \in \Pi(\overline{p}) \times \mathbb{R}$ (provided $u_1 = u_2$).

As shown in this theorem, more aversion to commitment implies $W_1(\pi, t) \geq W_2(\pi, t)$. That is, agent 1's net value of information is always greater than agent 2's. In other words, information acquisition is always more costly for agent 2. The behavioral comparison of Definition 4 has various implications regarding more specific representations.

Corollary 1 *Assume that $\succsim_1$ is more averse to commitment than $\succsim_2$.*

(1) If both preferences admit additive cost representations, then $c_1(\pi) \leq c_2(\pi)$ *for all* $\pi \in \Pi(\overline{p})$.

(2) If both preferences admit optimal waiting representations, then $\beta_1(\pi) \geq \beta_2(\pi)$ *and* $\gamma_1(\pi) \leq \gamma_2(\pi)$ *for all* $\pi \in \Pi(\overline{p})$.

(3) If both preferences admit constrained information representations, then for any $\pi_2 \in \Pi_2$, *there exists* $\pi_1 \in \Pi_1$ *with* $\pi_1 \trianglerighteq \pi_2$.

(4) If both preferences admit fixed information representations, then $\pi_1^* \trianglerighteq \pi_2^*$.

Proof. (1) An additive cost model is a special case where $W_i(\pi, t) = t - c_i(\pi)$. $W_1(\pi, t) \geq W_2(\pi, t)$ implies $c_1(\pi) \leq c_2(\pi)$.

(2) In the optimal waiting representation, $W_i(\pi, t) = \beta_i(\pi)t$ for $t > 0$. Then, $W_1(\pi, t) \geq W_2(\pi, t)$ implies $\beta_1(\pi) \geq \beta_2(\pi)$, while if $W_i(\pi, t) = \gamma_i(\pi)t$ for $t < 0$, $\gamma_1(\pi) \leq \gamma_2(\pi)$.

(3) If $W_1(\pi, t) \geq W_2(\pi, t)$, then $W_2(\pi, t) = t$ implies $W_1(\pi, t) = t$. Since a constrained information representation is a special case where $W(\pi, t) = t$ for all $\pi \in \Pi(\overline{p})$ such that $\pi' \trianglerighteq \pi$ for some $\pi' \in \Pi$, and $W(\pi, t) = -\infty$ otherwise, we have that for any $\pi_2 \in \Pi_2$, there exists $\pi_1 \in \Pi_1$ with $\pi_1 \trianglerighteq \pi_2$.

(4) This follows from part 3 since $\Pi_i = \{\pi_i^*\}$ in the fixed information representation. ■

Part (1) is shown by de Oliveira, Denti, Mihm, and Ozbek [15]. Part (4) is shown by Dillenberger, Lleras, Sadowski, and Takeoka [18].

3 Rationally inattentive representations

de Oliveira, Denti, Mihm, and Ozbek [15] axiomatize the additive cost representation for $\succsim$, which is called a rationally inattentive preference. We can characterize this class of representation as a special case of the Costly Subjective Learning Representation.

In addition to Singleton Independence and Aversion to Contingent Planning, de Oliveira, Denti, Mihm, and Ozbek [15] assume the following weakening of the Independence axiom:

Axiom 9 (Independence of Degenerate Decisions) *For all menus* F, G, *all acts* h, h', *and* $\alpha \in (0, 1)$,

$$\alpha F + (1 - \alpha)\{h\} \succsim \alpha G + (1 - \alpha)\{h\} \implies \alpha F + (1 - \alpha)\{h'\} \succsim \alpha G + (1 - \alpha)\{h'\}.$$

Note that this axiom is a further weakening of Certainty Independence: for all menus F, G, all acts h, and $\alpha \in (0, 1)$,

$$F \succsim G \iff \alpha F + (1 - \alpha)\{h\} \succsim \alpha G + (1 - \alpha)\{h\},$$

which is obviously weaker than Independence.

It is known that given the other axioms, Independence of Degenerate Decisions is equivalent to translation invariance: for all translations θ on X,[4]

$$F \succsim G \iff F + \theta \succsim G + \theta.$$

This property suggests that an incentive for costly information acquisition does not change under common translations, as discussed in Section 1.1.

As shown by de Oliveira, Denti, Mihm, and Ozbek [15, Theorem 1], an axiomatic foundation for the additive cost model is given as follows:[5]

Corollary 2 *Preference $\succsim$ satisfies the basic axioms, Singleton Independence, Aversion to Contingent Planning, and Independence of Degenerate Decisions if and only if it admits an additive cost representation, that is, there exists $(u, \overline{p}, c)$ such that*

$$U(F) = \max_{\pi \in \Pi(\overline{p})} \{b_F^u(\pi) - c(\pi)\}.$$

One observation is that together with quasi-convexity, translation invariance implies convexity of the representation. One may wonder what representation is characterized if quasi-convexity is strengthened to convexity without translation invariance being imposed. This class of preferences is weaker than the rationally inattentive preference, but is nested in the Costly Subjective Learning Representation.

We adapt an axiom of Mihm and Ozbek [28], which ensures convexity of the representation.

Axiom 10 (Increasing Desire for Commitment) *For any menus $F, G \in \mathbb{F}$ and lotteries $x, y \in X$, if $F \sim \{x\}$ and $G \sim \{y\}$, $\alpha\{x\} + (1-\alpha)\{y\} \succsim \alpha F + (1-\alpha)G$ for any $\alpha \in [0,1]$.*

The following result shows that if Aversion to Contingent Planning is replaced with Increasing Desire for Commitment, the Costly Subjective Learning Representation is forced to be convex, and then is reduce to the additive cost representation.

Corollary 3 *Preference $\succsim$ satisfies the basic axioms, Singleton Independence, and Increasing Desire for Commitment if and only if it admits an additive cost representation.*

This result suggests that if we want to retain convexity as a reasonable property of the representation, we cannot use any other model of subjective learning than the additive cost model. But, this result crucially depends on our assumption that u has an unbounded range. Since $V : C(\Delta(\Omega)) \to \mathbb{R}$ is monotone, normalized, convex, continuous, and the unbounded range $V(C(\Delta(\Omega))) = \mathbb{R}$, by a minor modification of Cerreia-Vioglio, Maccheroni, Marinacci,

[4] A translation θ is defined as $\theta = x - y$ for some $x, y \in X$. Accordingly, for all acts f, $f + \theta \in \mathcal{F}$ is defined as $f(\omega) + \theta$ for all ω as long as the operation is feasible. For all menus F, $F + \theta$ is the menu given by $\{f + \theta \mid f \in F\}$.

[5] In terms of the unboundedness axiom, they only require one-side of unboundedness.

and Montrucchio [8, Corollary 38] and Strzalecki [30, Theorem 3], $V : C(\Delta(\Omega)) \to \mathbb{R}$ is niveloid: for all $\varphi, \varphi' \in C(\Delta(\Omega))$,

$$V(\varphi) - V(\varphi') \leq \sup(\varphi - \varphi').$$

Since $V : C(\Delta(\Omega)) \to \mathbb{R}$ is monotone, normalized, and convex niveloid, de Oliveira, Denti, Mihm, and Ozbek [15, Theorem 1] imply that a preference $\succsim$ is represented by an additive cost representation.

4 Optimal waiting representations

In this section, we axiomatize a special case of the Costly Subjective Learning Representation where costs for information acquisition are measured as multiplicative waits on the value of information. As discussed in Section 1.1, such a specialization is suitable for modeling discounting costs for information acquisition.

4.1 Functional form

We consider the situation where a cost of choosing a signal structure π is measured by the waiting time or discount factor until a new signal arrives. Since the measure of cost for information acquisition changes between gains and losses, we introduce a reference point $x_0 \in X$ (such as the agent's initial endowment) whose utility is normalized to be zero.

Definition 5 *An Optimal Waiting Representation is a Costly Subjective Learning Representation* $(u, \overline{p}, W)$ *such that* $u(x_0) = 0$ *and for some* $\beta : \Pi(\overline{p}) \to [0, 1]$ *and* $\gamma : \Pi(\overline{p}) \to [1, \infty]$*,* W *is written as*

$$W(\pi, t) = \beta(\pi)(t)^+ - \gamma(\pi)(t)^-,$$

where $(t)^+ = \max\{0, t\}$ *and* $(t)^- = \max\{0, -t\}$ *for all* $t \in \mathbb{R}$*, and* $\infty \times 0 = 0$ *with convention.*

The function $\beta : \Pi(\overline{p}) \to [0, 1]$ is called a discounting function. The properties of the net benefit function W is inherited to β as follows:

(i) β is upper semi-continuous,

(ii) $\beta(\delta_{\overline{p}}) = 1$ for the initial prior $\overline{p}$,

(iii) for all $\pi, \rho \in \Pi(\overline{p})$, $\pi \trianglerighteq \rho \implies \beta(\pi) \leq \beta(\rho)$.

Part (i) is a technical condition to ensure a well-defined optimization problem of information acquisition. Part (ii) states that there is no cost (no discounting) if the prior information is chosen. Part (iii) states that if the agent wants to obtain a more informative signal structure, he has to wait longer, that is, its waiting cost corresponds to a smaller β.

The function $\gamma : \Pi(\overline{p}) \to [1, \infty]$ is called a premium function.[6] For some menu F, the agent has to end up with losses or negative values of information no matter what information structures are chosen. For example, suppose that if the agent makes a decision according to the prior, she immediately undertakes some losses. Alternatively, she postpones the payment of losses to conduct information acquisition. The investment decision can be made according to more informative signals and presumably the losses are reduced. The trade-off is that the losses have to be paid back with some premium. In such a situation, costs for choosing information structure are measured by the premium function. The properties of the net benefit function W is inherited to γ as follows:

(i) γ is lower semi-continuous,

(ii) $\gamma(\delta_{\overline{p}}) = 1$ for the initial prior $\overline{p}$,

(iii) for all $\pi, \rho \in \Pi(\overline{p})$, $\pi \trianglerighteq \rho \implies \gamma(\pi) \geq \gamma(\rho)$.

Parts (i) and (ii) are symmetric to those of discounting function. Part (iii) states that if the agent wants to obtain a more informative signal structure, he has to pay more premium.

To interpret the functional form in Definition 5, define

$$\Pi^+(F) = \{\pi \in \Pi(\overline{p}) \,|\, b_F^u(\pi) \geq 0\}.$$

Note that $\Pi^+(F) \neq \emptyset$ if $b_F^u(\pi) \geq 0$ for some π, and $\Pi^+(F) = \emptyset$ if $b_F^u(\pi) < 0$ for all π. Then, the above functional form can be rewritten as

$$U(F) = \begin{cases} \max_{\pi\in\Pi} \beta(\pi) b_F^u(\pi) \geq 0 & \text{if } \Pi^+(F) \neq \emptyset, \\ \max_{\pi\in\Pi} \gamma(\pi) b_F^u(\pi) < 0 & \text{if } \Pi^+(F) = \emptyset. \end{cases} \tag{7}$$

In this representation, the agent behaves as if he optimally chooses an information structure π by taking into account the associated waiting time (captured by discount function $\beta(\pi)$ or premium function $\gamma(\pi)$). The benefit of a signal π is its information value $b_F^u(\pi)$. The agent optimally chooses a signal that maximizes benefits multiplied by costs associated with waiting time.

On the singleton menus, U coincides with

$$\begin{aligned} U(F) &= \begin{cases} \max_{\pi\in\Pi} \beta(\pi) \sum_\Omega u(f(\omega)) p^\pi(\omega) & \text{if } \Pi^+(\{f\}) \neq \emptyset, \\ \max_{\pi\in\Pi} \gamma(\pi) \sum_\Omega u(f(\omega)) p^\pi(\omega) & \text{if } \Pi^+(\{f\}) = \emptyset \end{cases} \\ &= \sum_\Omega u(f(\omega)) \overline{p}(\omega). \end{aligned}$$

Since there is no role of information acquisition under commitment, the decision maker just sticks to his initial belief $\overline{p}$.

[6] An information structure π with $\gamma(\pi) = \infty$ is not relevant since it is too costly and never chosen in the stage of information acquisition.

Figure 1 summarizes the timing of decisions. Given a menu F, suppose that the agent chooses an information structure, say, π. After some waiting time, he receives an signal and updates the initial belief $\overline{p}$ to p. The agent chooses an act $f \in F$ by maximizing the subjective expected utility $\sum u(f(\omega))p(\omega)$. Afterwards, a state ω is realized and the agent receives a lottery $f(\omega)$. Finally, he receives a prize a according to the realization of $f(\omega)$. Alternatively, the agent may choose a more informative signal structure with longer waiting time, and make a choice from F according to the realization of signals. The agent optimally solves this costly information acquisition menu by menu.

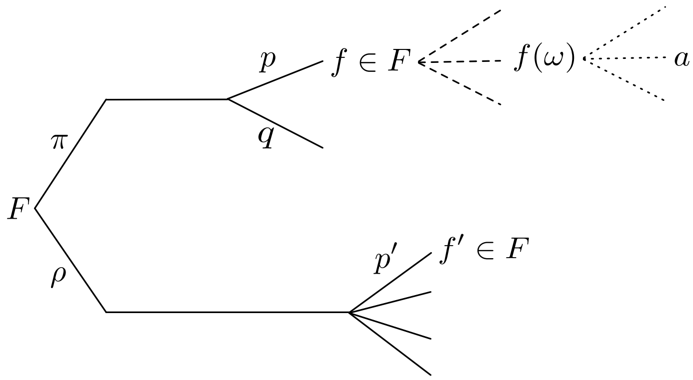

Figure 1: Timing of decisions

4.1.1 Single-crossing property under translations

Consider two information structures such that one is more informative than the other. When the agent makes a choice on information acquisition, he has to solve a trade-off between higher value of information and longer waiting time. As discussed in Section 1.1, we illustrate how the willingness to wait for information is affected by adding a constant payoff to menus.

Fix any lottery $x \in X$. For any utility level $v \in \mathbb{R}$, define a v-translation of x as a lottery x^v satisfying $u(x^v) = u(x) + v$. If u is unbounded, there exists some v-translation of x. For any act $f \in \mathcal{F}$, a v-translation of f is defined as an act obtained by applying the above translation state by state. Finally, for any menu F, a v-translation of F is defined as a menu obtained by applying the above translation element by element. For each menu F, its v-translation is denoted by $F + v$.

Fix a menu F arbitrarily. We denote the net value of information π under $F + v$ by

$$W_{F,\pi}(v) = \beta(\pi)(b^u_{F+v}(\pi))^+ - \gamma(\pi)(b^u_{F+v}(\pi))^-.$$

Take two information structures π^H and π^L with $\pi^H \trianglerighteq \pi^L$. By assumption, $b^u_F(\pi^H) \geq b^u_F(\pi^L)$. To separate cases of gains and losses, let v^H and v^L be translations under which the gross values of information of π^H and π^L are zero, respectively. That is, $b^u_{F+v^H}(\pi^H) = b^u_{F+v^L}(\pi^L) = 0$. Since $b^u_{F+v}(\pi) = b^u_F(\pi) + v$, we have $v^L = -b^u_F(\pi^L) \geq -b^u_F(\pi^H) = v^H$.

Hence, for any $v > v^L$, the menu $F + v$ gives positive values of information, $b^u_{F+v}(\pi^H) \geq b^u_{F+v}(\pi^L) > 0$, while for any $v < v^H$, the menu $F + v$ gives negative values of information, $b^u_{F+v}(\pi^L) \leq b^u_{F+v}(\pi^H) < 0$.

The following proposition shows the single crossing property of $W_{F,\pi^H}(v)$ and $W_{F,\pi^L}(v)$ on each domain of gains and losses.

Proposition 1 *Assume that $\pi^H \trianglerighteq \pi^L$ with $\beta(\pi^H) > 0$ and $\gamma(\pi^H) < \infty$. The following statements hold:*

(i) For $v' \geq v > v^L$, we have

$$W_{F,\pi^L}(v) \geq (>) W_{F,\pi^H}(v) \Rightarrow W_{F,\pi^L}(v') \geq (>) W_{F,\pi^H}(v').$$

(ii) For $v^L \geq v \geq v^H$, we have

$$W_{F,\pi^H}(v) \geq 0 \geq W_{F,\pi^L}(v).$$

(iii) For $v^H > v' \geq v$, we have

$$W_{F,\pi^L}(v') \geq (>) W_{F,\pi^H}(v') \Rightarrow W_{F,\pi^L}(v) \geq (>) W_{F,\pi^H}(v).$$

Part (i) states that if the agent is reluctant to choose a more informative signal structure at a moderate payoff translation v, so is he at a greater payoff translation v'. Similarly, part (iii) states that if the agent is reluctant to reduce losses by choosing a more informative signal structure at a moderate v, so is he when the menu is translated further toward the direction of losses. This result suggests that the agent has a weaker incentive for acquiring additional information when a constant payoff with a greater magnitude is added to menus on each domain of gains and losses.

4.2 Axiom

We turn to the behavioral foundation of the optimal waiting representation. As shown in the previous subsection, adding constants affects the incentive for information acquisition, which may lead to violation of Independence. For its axiomatization, we identify instances where the incentive for information acquisition is invariant.

Suppose that information acquisition entails a time-consuming task and the agent cares about payoff-discounting from waiting time until arrival of signals. Suppose that payoffs are given as utils and $\{(100, 0), (0, 100)\} \sim \{(60, 60)\}$. For a positive α, the menu $\{(\alpha 100, 0), (0, \alpha 100)\}$ is interpreted as a scale up or down of the menu $\{(100, 0), (0, 100)\}$ toward the origin $(0, 0)$. Since α is a common multiplier for all payoffs, it is independent of discounting, and hence does not affect an incentive for information acquisition. Consequently, the agent will be indifferent between $\{(\alpha 100, 0), (0, \alpha 100)\}$ and $\{(\alpha 60, \alpha 60)\}$.

The following axiom is motivated by the above observation. Recall that x_0 is a reference point.

Axiom 11 (Reference-Point Independence) *For all menus F and G and $\alpha \in (0,1)$,*

$$F \succsim G \iff \alpha F + (1-\alpha)\{x_0\} \succsim \alpha G + (1-\alpha)\{x_0\}.$$

This axiom states that Independence holds only when menus are mixed with the reference outcome.

Theorem 4 *Preference $\succsim$ satisfies the basic axioms, Singleton Independence, Aversion to Contingent Planning, and Reference-Point Independence if and only if it admits an optimal waiting representation, that is, there exists $(u, \overline{p}, \beta, \gamma)$ such that*

$$U(F) = \max_{\pi \in \Pi(\overline{p})} [\beta(\pi)(b_F^u(\pi))^+ - \gamma(\pi)(b_F^u(\pi))^-].$$

Compared with de Oliveira, Denti, Mihm, and Ozbek [15], we can characterize the optimal waiting representation by replacing Independence of Degenerate Decisions with Reference-Point Independence.

The following is a proof sketch of the sufficiency. The axioms imply that $\succsim$ admits a Costly Subjective Learning Representation such as

$$V(\varphi) = \max_{\pi \in \Delta(\Delta(\Omega))} W(\pi, \langle \varphi, \pi \rangle),$$

where φ is a continuous function on $\Delta(\Omega)$ and $\langle \varphi, \pi \rangle = \int \varphi(p) \, \mathrm{d}\pi(p)$. Since $u : X \to \mathbb{R}$ is an expected utility, we can assume $u(x_0) = 0$, where x_0 is a reference point. Since $\succsim$ satisfies Reference-Point Independence, we can show that V is positively homogeneous, which in turn implies that $W(\pi, t)$ is homogeneous of degree one in t. By defining $\beta(\pi) = W(\pi, 1)$ and $\gamma(\pi) = -W(\pi, -1)$, V can be rewritten as

$$V(\varphi) = \max_{\pi \in \Delta(\Delta(\Omega))} [\beta(\pi)(\langle \varphi, \pi \rangle)^+ - \gamma(\pi)(\langle \varphi, \pi \rangle)^-].$$

This type of specification of W has a counterpart in decision making under ambiguity, which is called the confidence model (Chateauneuf and Faro [10]). Finally, we show that discounting and premium functions, β and γ, satisfy the desired properties.

4.3 Uniqueness

As a preliminary, we start with the uniqueness of risk preference u and a prior belief $\overline{p}$.

Proposition 2 *Assume that there are two optimal waiting representations $(u_i, \overline{p}_i, \beta_i, \gamma_i)$, $i = 1, 2$ that represent the same preference $\succsim$ on $\mathbb{F}$. Then, there exists $\alpha > 0$ such that $u_2 = \alpha u_1$ and $\overline{p}_1 = \overline{p}_2$.*

Since $\succsim$ is represented by a subjective expected utility, this uniqueness directly follows from the uniqueness of Anscombe and Aumann [2]. Since we assume that $u_i(x_0) = 0$ for all i, we have $u_2 = \alpha u_1$.

We turn to the uniqueness of discounting function β. From now on, we assume that we have already pinned down $(u, \overline{p})$ uniquely. We start with a casual observation. For all π and all menus F, the optimal waiting representation implies that for all F and π with $b_F^u(\pi) > 0$,

$$U(F) = \max_{\rho \in \Pi} \beta(\rho) b_F^u(\rho) \geq \beta(\pi) b_F^u(\pi).$$

Thus, we have, for all such a π,

$$\beta(\pi) \leq \frac{U(F)}{b_F^u(\pi)}.$$

If π is an optimal information structure at F,

$$\beta(\pi) = \frac{U(F)}{b_F^u(\pi)}.$$

We extend this formula to $\Pi(\overline{p})$ and define a canonical discount function as follows:

Definition 6 $\beta^* : \Pi(\overline{p}) \to [0, 1]$ *is said to be a canonical discounting function if*

$$\beta^*(\pi) = \inf_{\{F \in \mathbb{F} \,|\, b_F^u(\pi) > 0\}} \frac{u(x_F)}{b_F^u(\pi)},$$

where $x_F \in X$ *is a lottery equivalent of* F*, that is,* $\{x_F\} \sim F$*.*

Since $U(F) = u(x_F)$, this definition is consistent with the above argument. Note that β^* is invariant for all $u' = \alpha u$ from Proposition 2. Thus, β^* is uniquely determined from preference and is interpreted as a measure of discounting costs from information acquisition.

To identify the premium function, take any F such that $b_F^u(\pi) < 0$ for all π. According to the representation,

$$U(F) = \max_{\rho \in \Pi} \gamma(\rho) b_F^u(\rho) \geq \gamma(\pi) b_F^u(\pi).$$

Since $b_F^u(\pi) < 0$,

$$\gamma(\pi) \geq \frac{U(F)}{b_F^u(\pi)},$$

where the equality holds if π is an optimal information structure at F. Thus,

$$\gamma(\pi) = \sup_{\{F \in \mathbb{F} \,|\, b_F^u(\pi) < 0\}} \frac{u(x_F)}{b_F^u(\pi)}.$$

Definition 7 $\gamma^* : \Pi(\overline{p}) \to [1, \infty]$ *is said to be a canonical premium function if*

$$\gamma^*(\pi) = \sup_{\{F \in \mathbb{F} \,|\, b_F^u(\pi) < 0\}} \frac{u(x_F)}{b_F^u(\pi)}.$$

By Proposition 2, γ^* is invariant for all positive affine transformations of u.

Say that an optimal waiting representation $(u, \overline{p}, \beta, \gamma)$ is canonical if both β and γ are canonical.

Theorem 5 *If $\succsim$ satisfies the axioms of Theorem 4, then it admits a canonical optimal waiting representation $(u, \overline{p}, \beta^*, \gamma^*)$. If there exist two canonical representations, $(u_i, \overline{p}_i, \beta_i^*, \gamma_i^*)$, $i = 1, 2$, then there exists $\alpha > 0$ such that $u_2 = \alpha u_1$, $\overline{p}_1 = \overline{p}_2$, $\beta_1^* = \beta_2^*$, and $\gamma_1^* = \gamma_2^*$.*

The expressions of β^* and γ^* given as in Definitions 6 and 7 provide an explicit formula for eliciting discounting and premium functions. If u and $\{x_F\}$ are elicited from the agent's preference, $\beta^*(\pi)$ and $\gamma^*(\pi)$ can be computed according to their definitions.

Our uniqueness result has a parallel relationship with de Oliveira, Denti, Mihm, and Ozbek [15]. They define a canonical additive cost function on $\Pi(\overline{p})$ by the formula,

$$c^*(\pi) = \sup_{F \in \mathbb{F}} (b_F^u(\pi) - u(x_F)),$$

whereby a unique canonical representation is obtained as

$$U(F) = \max_{\pi \in \Pi(\overline{p})} (b_F^u(\pi) - c^*(\pi)).$$

4.4 Application: Optimal sampling under discounting costs

Cukierman [11] investigate an optimal number of information acquisition periods before an investment decision is made with assuming additive costs for information acquisition. We adopt the same setting with assuming discounting costs and see its implications.

The state space Ω is taken to be the real line. The prior over Ω is given by a normal distribution $\omega \sim N(\mu, 1/\tau)$, where μ is the mean and $\tau > 0$ is the precision. The signal s is correlated with ω according to a normal distribution $s \sim N(\omega, 1/\sigma)$, where $\sigma > 0$ is the precision of the signal.

The agent's payoff function is state-dependent and given by

$$u(y, \omega) = a\omega - b|\omega - y|, \ a > 0, \ b > 0.$$

A choice variable is y, which is interpreted as an investment decision. This payoff function takes its maximum at $y = \omega$. Since payoffs change according to realization of ω, a choice of y is interpreted as a choice of act.

The agent can postpone the investment decision and instead observe signals, whereby the prior is updated to a posterior according to Bayes' rule. If signals are observed for t periods, the value of information is given by

$$b^u(t) = \int \max_y \int u(y, \omega) \, \mathrm{d}p(\omega | s_1, \cdots, s_t) \, \mathrm{d}\pi^t(s_1, \cdots, s_t), \tag{8}$$

where $p(\omega | s_1, \cdots, s_t)$ is a posterior conditional upon the realization of signals $s_1, \cdots, s_t$ and $\pi^t(s_1, \cdots, s_t)$ is an ex ante probability of the signal realization up to period t. In this setting, an information structure is identified with a number of periods for signal observations. A more informative signal structure is obtained by waiting longer. The set of information structures is given by $\Pi = \{\pi^t \mid t \geq 0\}$.

Assume that the agent's preference is represented by an optimal waiting representation such that a discounting function is given by an exponential discounting $\beta(\pi^t) = e^{-\gamma t}$ with a constant $\gamma > 0$. The agent solves an optimal sampling problem formulated as

$$\max_t e^{-\gamma t} b^u(t).$$

Cukierman [11] shows that (8) is written as

$$b^u(t) = a\mu - b\left(\frac{2}{\pi}\right)^{\frac{1}{2}}\left(\frac{1}{\tau + t\sigma}\right)^{\frac{1}{2}},$$

where π is the circular constant. From the FOC of $\max_t e^{-\gamma t} b^u(t)$,

$$\frac{\frac{\mathrm{d}b^u}{\mathrm{d}t}(t)}{b^u(t)} = \gamma. \tag{9}$$

To ensure that $b^u(t) > 0$ for all t, assume that $a\mu - b\left(\frac{2}{\pi}\right)^{\frac{1}{2}} > 0$. Then, it is easy to see that if $b^u(t) > 0$ and $\frac{\mathrm{d}^2 b^u}{\mathrm{d}t^2}(t) < 0$, then $\frac{\frac{\mathrm{d}b^u}{\mathrm{d}t}(t)}{b^u(t)}$ is strictly decreasing. Thus, the SOC is satisfied. Therefore, (9) is a necessary and sufficient condition for an optimal waiting time t.

Now some comparative statics are possible. Note that the left-hand side of (9) is explicitly written as

$$\frac{b\sigma\left(\frac{2}{\pi}\right)^{\frac{1}{2}}\left(\frac{1}{\tau+t\sigma}\right)^{\frac{3}{2}}}{a\mu - b\left(\frac{2}{\pi}\right)^{\frac{1}{2}}\left(\frac{1}{\tau+t\sigma}\right)^{\frac{1}{2}}}$$

or

$$f(x) = \frac{b\sigma\left(\frac{2}{\pi}\right)^{\frac{1}{2}} x^{\frac{3}{2}}}{a\mu - b\left(\frac{2}{\pi}\right)^{\frac{1}{2}} x^{\frac{1}{2}}}, \text{ where } x = \frac{1}{\tau + t\sigma}.$$

On the region where $a\mu - b\left(\frac{2}{\pi}\right)^{\frac{1}{2}} x^{\frac{1}{2}} > 0$, $f(x)$ is strictly increasing. If the precision τ of the prior decreases, that is, the agent becomes more uncertain about the states, $f(\frac{1}{\tau+t\sigma})$ moves upwards. Thus, the agent will wait longer for information acquisition.

If the mean μ of the prior increases, that is, the investment becomes more profitable on average, then f moves downwards. Thus, the agent will quit information acquisition earlier.

These comparative statics are summarized in the following proposition.

Proposition 3 *Assume the waiting cost model and $a\mu - b\left(\frac{2}{\pi}\right)^{\frac{1}{2}} > 0$. The agent waits longer for signal observations if either the variance of the prior over states increases or the mean of the prior over states decreases.*

Cukierman [11] considers the same problem by assuming the additive cost for information acquisition. The agent solves

$$\max_t \{b^u(t) - ct\},$$

where $c > 0$ is a constant marginal cost of sampling. Then, the FOC is given by

$$\frac{\mathrm{d}b^u}{\mathrm{d}t}(t) = b\sigma \left(\frac{2}{\pi}\right)^{\frac{1}{2}} \left(\frac{1}{\tau + t\sigma}\right)^{\frac{3}{2}} = c.$$

It is easy to see that if τ increases, $\frac{\mathrm{d}b^u}{\mathrm{d}t}(t)$ shifts up, and hence, an optimal waiting time increases. This observation is the same as in Proposition 3. On the other hand, the FOC in the additive cost model is independent of μ, which implies that the mean of the prior has no impact on the optimal waiting time. This has a clear contrast with the implications of waiting time model.

5 Constrained information representations

In this section, we consider special cases of the Costly Subjective Learning Representation, which are nested in both additive cost and optimal waiting representations.

As mentioned in Section 2.3, the constrained information representation is the model where the agent chooses an optimal information structure from the set of subjectively possible information structure without any costs. This class of representations is characterized by imposing both Independence of Degenerate Decisions and Reference-Point Independence. Formally, the result is given as follows:

Theorem 6 *Assume that preference $\succsim$ satisfies the basic axioms, Singleton Independence, Aversion to Contingent Planning, and Reference-Point Independence. Then, $\succsim$ satisfies Independence of Degenerate Decisions if and only if it admits a constrained information representation, that is, there exists $(u, \overline{p}, \Pi)$ such that*

$$U(F) = \max_{\pi \in \Pi} b^u_F(\pi).$$

In the proof of sufficiency, we show that if $\succsim$ satisfies Independence of Degenerate Decisions in addition, the optimal waiting representation is rewritten as a constrained information representation. de Oliveira, Denti, Mihm, and Ozbek [15] show a symmetric result. They show that if $\succsim$ satisfies Reference-Point Independence in addition, the additive costs representation is rewritten as a constrained information representation.

A further special case is the fixed information representation, which is investigated in Dillenberger, Lleras, Sadowski, and Takeoka [18].

Corollary 4 *Preference $\succsim$ satisfies the basic axioms and Independence if and only if it admits a fixed information representation, that is, there exists $(u, \overline{p}, \Pi = \{\pi^*\})$ such that*

$$U(F) = b^u_F(\pi^*).$$

Since Independence implies all the axioms of Theorem 6, $\succsim$ admits a constrained information representation with Π. In the proof of sufficiency, we show that if $\succsim$ satisfies Independence, Π is reduced to a singleton set.

Appendix

A Preliminaries

Following de Oliveira, Denti, Mihm, and Ozbek [15], we introduce some notions and mathematical preliminaries needed for the subsequent analysis. The proofs are omitted.

- $C(\Delta(\Omega))$: the set of all real-valued continuous functions over $\Delta(\Omega)$ with the supnorm
- $ca(\Delta(\Omega))$: the set of all signed measures over $\Delta(\Omega)$ with the weak* topology
- $ca_+(\Delta(\Omega))$: the set of all positive measures over $\Delta(\Omega)$
- For $\varphi \in C(\Delta(\Omega))$ and $\pi \in ca(\Delta(\Omega))$, define

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

For a subset Ψ of $C(\Delta(\Omega))$, we say that a function $V : \Psi \to \mathbb{R}$ is *normalized* if $V(\alpha) = \alpha$ for each constant function $\alpha \in \Psi$; *monotone* if $V(\varphi) \geq V(\psi)$ for all $\varphi, \psi \in \Psi$; *convex* if $\alpha V(\varphi) + (1-\alpha)V(\psi) \geq V(\alpha\varphi + (1-\alpha)\psi)$ for all $\varphi, \psi \in \Psi$ and $\alpha \in (0,1)$; *quasi-convex* if $V(\varphi) \geq V(\alpha\varphi + (1-\alpha)\psi)$ for all $\varphi, \psi \in \Psi$ with $V(\varphi) \geq V(\psi)$ and $\alpha \in (0,1)$; *positively homogeneous* if $V(\alpha\varphi) = \alpha V(\varphi)$ for all $\varphi \in \Psi$ and $\alpha \geq 0$.

- Φ: the set of convex functions in $C(\Delta(\Omega))$
- Φ^*: the dual cone of Φ, that is,

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

  The set Φ^* is also a closed convex cone such that $0 \in \Phi^*$.

- For any expected utility function u and any menu $F \in \mathbb{F}$, let

$$\varphi_F(p) = \max_{f \in F} \sum_{\Omega} u(f(\omega))p(\omega)$$

- $\Phi_{\mathbb{F}}(\Phi_{\mathcal{F}}, \Phi_X)$: the set of functions $\varphi_F(\varphi_{\{f\}}, \varphi_{\{x\}})$

Note that $u(X) = \Phi_X \subset \Phi_{\mathcal{F}} \subset \Phi_{\mathbb{F}} \subset \Phi$. Moreover, $\Phi_{\mathbb{F}}$ is convex because $\alpha\varphi_F + (1-\alpha)\varphi_G = \varphi_{\alpha F+(1-\alpha)G}$.

Assume that $u(X) = \mathbb{R}$. Then we have the following properties of $\Phi_{\mathbb{F}}$:

(i) $\Phi_{\mathbb{F}} \subset \Phi$

(ii) $\Phi_{\mathbb{F}} + \mathbb{R} = \Phi_{\mathbb{F}}$

(iii) $\alpha\varphi_F \in \Phi_{\mathbb{F}}$ for every $\alpha \geq 0$

(iv) The set $\Phi_{\mathbb{F}}$ is dense in Φ.

B Proof of Theorem 1

B.1 Sufficiency

First, we derive a utility representation $U : \mathbb{F} \to \mathbb{R}$ and define the functional $V : \Phi_{\mathbb{F}} \to \mathbb{R}$ as in de Oliveira, Denti, Mihm, and Ozbek [15].

Claim 1 *Every menu F has a certainty equivalent $x_F \in X$ such that $x_F \sim F$.*

Claim 2 *There exists an expected utility function $u : X \to \mathbb{R}$ with unbounded range and a prior probability measure $\overline{p}$ over Ω such that the preference $\succsim$ over $\mathcal{F}$ is represented by the function $U : \mathcal{F} \to \mathbb{R}$ defined by*

$$U(f) = \sum_{\Omega} u(f(\omega))\overline{p}(\omega)$$

We extend $U : \mathcal{F} \to \mathbb{R}$ to $\mathbb{F}$ by $U(F) = U(x_F)$. By claim 1, $U : \mathbb{F} \to \mathbb{R}$ represents $\succsim$. Without loss of generality, we can assume $u(x_0) = 0$, and by Unboundedness, $U(\mathbb{F}) = \mathbb{R}$.

Define the functional $V : \Phi_{\mathbb{F}} \to \mathbb{R}$ by $V(\varphi_F) = U(F)$. as in de Oliveira, Denti, Mihm, and Ozbek [15]. They show that V is well-defined.

Lemma 1 *The functional $V : \Phi_{\mathbb{F}} \to \mathbb{R}$ is monotone, normalized, quasi-convex, and continuous.*

Proof. The first two properties follow from the same argument of Claim 6 in de Oliveira, Denti, Mihm, and Ozbek [15].

Claim 3 $V : \Phi_{\mathbb{F}} \to \mathbb{R}$ *is quasi-convex.*

Proof. Following Cerreia-Vioglio, Maccheroni, Marinacci, and Montrucchio [8, Lemma 56], quasi-convexity of V is obtained as follows: We want to show that for all F, G, $U(F) \geq U(G)$ implies $U(F) \geq U(\alpha G + (1-\alpha)F)$ for all $\alpha \in (0,1)$. If $F \sim G$, the desired result directly follows from Aversion to Contingent Planning. Then, we show that $F \succ G$ implies that $F \succsim \alpha G + (1-\alpha)F$ for all $\alpha \in (0,1)$. Suppose contrary that there exist $F \succ G$ and $\widetilde{\alpha} \in (0,1)$ such that $\widetilde{\alpha}G + (1-\widetilde{\alpha})F \succ F$. Note that $\widetilde{\alpha} \in \{\alpha \in [0,1] | \, \alpha G + (1-\alpha)F \succsim F\} \neq \emptyset$. By Mixture Continuity, this set is compact. Hence, we can find $\beta = \max\{\alpha \in [0,1] | \, \alpha G + (1-\alpha)F \succsim F\}$ and define $F_\beta = \beta G + (1-\beta)F$.

We claim that $F_\beta \sim F$. If $\beta = 1$, then $G \succsim F$, which contradicts $F \succ G$. Hence, $\beta < 1$. Now we show that $F_\beta \sim F$. Suppose contrary that $F_\beta \nsim F$, that is, $F \succ F_\beta$. Since $\{\alpha \in [0,1] | \, \alpha G + (1-\alpha)F \succ F\}$ is open, we can find an open set V such that $\beta \in V$ and $V \subset \{\alpha \in [0,1] | \, \alpha G + (1-\alpha)F \succ F\}$. Hence there exists $\beta' \in V$ such that $\beta' \in (\beta, 1)$ and $\beta' G + (1-\beta')F \succ F$. This contradicts the maximality of β. Hence, $F_\beta \sim F$.

Since $F_\beta \sim F$, Aversion to Contingent Planning implies that $F \succsim \lambda F_\beta + (1-\lambda)F$ for all $\lambda \in (0,1)$. Since $0 < \widetilde{\alpha} < \beta$, $\frac{\widetilde{\alpha}}{\beta} \in (0,1)$. Thus, $F \succsim \frac{\widetilde{\alpha}}{\beta}F_\beta + (1-\frac{\widetilde{\alpha}}{\beta})F = \frac{\widetilde{\alpha}}{\beta}[\beta G + (1-\beta)F] + (1-\frac{\widetilde{\alpha}}{\beta})F = \widetilde{\alpha}G + (1-\widetilde{\alpha})F \succ F$, which is a contradiction. Hence, $F \succ G$ implies $F \succsim \alpha G + (1-\alpha)F$ for all $\alpha \in (0,1)$, as desired. ■

Let $\|\cdot\|$ be sup-norm. If $\{\varphi_n\}$ is a sequence in $\Phi_{\mathbb{F}}$, we write $\varphi_n \searrow \varphi$ if it is decreasing and it converges to φ in norm. The function $V : \Phi_{\mathbb{F}} \to \mathbb{R}$ is right continuous at $\varphi \in \Phi_{\mathbb{F}}$ if $\{\varphi_n\}_n \subseteq \Phi_{\mathbb{F}}$ and $\varphi_n \searrow \varphi$ implies $V(\varphi_n) \to V(\varphi)$. The function $V : \Phi_{\mathbb{F}} \to \mathbb{R}$ is upper semi-continuous if for any $\lambda > V(\varphi_F)$, there exists $\varepsilon > 0$ such that $\lambda > V(\varphi_{F'})$ for any $\varphi_{F'} \in \Phi_{\mathbb{F}}$ with $\|\varphi_F - \varphi_{F'}\| < \varepsilon$.

Similarly, if $\{\varphi_n\}$ is a sequence in $\Phi_{\mathbb{F}}$, we write $\varphi_n \nearrow \varphi$ if it is increasing and it converges to φ in norm. The function $V : \Phi_{\mathbb{F}} \to \mathbb{R}$ is left continuous at $\varphi \in \Phi_{\mathbb{F}}$ if $\{\varphi_n\}_n \subseteq \Phi_{\mathbb{F}}$ and $\varphi_n \nearrow \varphi$ implies $V(\varphi_n) \to V(\varphi)$. The function $V : \Phi_{\mathbb{F}} \to \mathbb{R}$ is lower semi-continuous if for any $\lambda < V(\varphi_F)$, there exists $\varepsilon > 0$ such that $\lambda < V(\varphi_{F'})$ for any $\varphi_{F'} \in \Phi_{\mathbb{F}}$ with $\|\varphi_F - \varphi_{F'}\| < \varepsilon$. The function V is continuous if it is upper and lower semi-continuous.

Claim 4 $V : \Phi_{\mathbb{F}} \to \mathbb{R}$ *is continuous.*

Proof. We show the claim in a similar way to Cerreia-Vioglio, Maccheroni, Marinacci, and Montrucchio [8, Lemma 42]. Note that $\Phi_{\mathbb{F}}$ is convex.

First, we show that V is upper semi-continuous.

Step 1: For any $F, G, H \in \mathbb{F}$, Mixture Continuity implies that the following sets are closed:

$$\begin{aligned}
\{\alpha \in [0,1] \,|\, \alpha F + (1-\alpha)G \succsim H\} &= \{\alpha \in [0,1] \,|\, U(\alpha F + (1-\alpha)G) \geq U(H)\} \\
&= \left\{\alpha \in [0,1] \,|\, V(\varphi_{\alpha F + (1-\alpha)G}) \geq V(\varphi_H)\right\} \\
&= \{\alpha \in [0,1] \,|\, V(\alpha\varphi_F + (1-\alpha)\varphi_G) \geq V(\varphi_H)\} \\
&= \{\alpha \in [0,1] \,|\, V(\alpha\varphi_F + (1-\alpha)\varphi_G) \geq \lambda\},
\end{aligned}$$

where $\lambda = V(\varphi_H)$.

Step 2: For any $\lambda \in \mathbb{R}$ and $\varphi, \varphi' \in \Phi_{\mathbb{F}}$ with $\varphi' \geq \varphi$ and $V(\varphi) < \lambda$, there exists $\alpha \in (0,1)$ such that $V(\alpha\varphi + (1-\alpha)\varphi') < \lambda$. Take such φ, φ', and λ. Suppose contrary that $V(\alpha\varphi + (1-\alpha)\varphi') \geq \lambda$ for all $\alpha \in (0,1)$. By Step 1, the set $A = \{\alpha \in [0,1] | V(\alpha\varphi + (1-\alpha)\varphi') \geq \lambda\}$ is closed. Since $(0,1) \subseteq A$, we have that $A = [0,1]$. This implies that $V(\varphi) \geq \lambda$, which is a contradiction.

Step 3: V is right continuous. Let $\varphi_n \searrow \overline{\varphi}$ such that $\{\varphi_n\}_{n\in\mathbb{N}} \cup \{\overline{\varphi}\} \subseteq \Phi_{\mathbb{F}}$. Monotonicity implies that $V(\varphi_n) \geq V(\varphi_{n+1}) \geq V(\overline{\varphi})$ for all $n \in \mathbb{N}$. Suppose contrary that $V(\varphi_n)$ does not converges to $V(\overline{\varphi})$, that is, there exists $\lambda \in \mathbb{R}$ such that $V(\varphi_n) \geq \lambda > V(\overline{\varphi})$ for all $n \in \mathbb{N}$. By Step 2, for each $\varphi \in \Phi_{\mathbb{F}}$ with $\varphi \geq \overline{\varphi}$, there exists $\alpha \in (0,1)$ such that $V((1-\alpha)\overline{\varphi} + \alpha\varphi) < \lambda$. Take $\varepsilon > 0$ such that $\overline{\varphi} + \varepsilon\mathbf{1} \in \Phi_{\mathbb{F}}$. Define $\varphi = \overline{\varphi} + \varepsilon\mathbf{1}$ and note that $\Phi_{\mathbb{F}} \ni (1-\alpha)\overline{\varphi} + \alpha\varphi = \overline{\varphi} - \alpha\overline{\varphi} + \alpha\overline{\varphi} + \alpha\varepsilon\mathbf{1} = \overline{\varphi} + \alpha\varepsilon\mathbf{1}$. Since $\varphi_n \searrow \overline{\varphi}$, there exists $\overline{n} \in \mathbb{N}$ such that $\varphi_n \leq \overline{\varphi} + \alpha\varepsilon\mathbf{1} = (1-\alpha)\overline{\varphi} + \alpha\varphi$ for all $n \geq \overline{n}$. Monotonicity implies that $V(\varphi_n) \leq V((1-\alpha)\overline{\varphi} + \alpha\varphi) < \lambda$ for all $n \geq \overline{n}$, which is a contradiction.

Step 4: The result. Let $\lambda \in \mathbb{R}$ and $S(V,\lambda) = \{\varphi \in \Phi_{\mathbb{F}} | V(\varphi_F) \geq \lambda\}$. We show that $\{\varphi_n\}_{n\in\mathbb{N}} \subseteq S(V,\lambda)$ and $\varphi_n \to \varphi \in \Phi_{\mathbb{F}}$ imply $\varphi \in S(V,\lambda)$. There exists $\overline{\varepsilon} > 0$ such that $\varphi + \varepsilon\mathbf{1} \in \Phi_{\mathbb{F}}$ for all $\varepsilon \in [0, \overline{\varepsilon}]$. Let $\varepsilon_m > 0$ be such that $\{\varepsilon_m\}_{m\in\mathbb{N}} \subseteq [0,\overline{\varepsilon}]$ and $\varepsilon_m \searrow 0$. Note that $\varphi + \varepsilon_m\mathbf{1} \in \Phi_{\mathbb{F}}$ for all $m \in \mathbb{N}$. Since $\varphi_n \to \varphi$, for all $m \in \mathbb{N}$ there exists n_m such that $\varphi + \varepsilon_m\mathbf{1} \geq \varphi_{n_m}$. Monotonicity implies that $V(\varphi + \varepsilon_m\mathbf{1}) \geq V(\varphi_{n_m}) \geq \lambda$. By right continuity, we have that $V(\varphi) = \lim_m V(\varphi + \varepsilon_m\mathbf{1}) \geq \lambda$.

Notice that the above proof goes through when $\succsim$ satisfies Mixture Continuity and V is monotone. Hence, by the symmetric argument, we can show that V is also lower semi-continuous. ■

■

Define an extension of V to $C(\Delta(\Omega))$ by

$$V(\varphi) = \inf\{V(\varphi_F) | \varphi_F \in \Phi_{\mathbb{F}}, \varphi_F \geq \varphi\} \tag{10}$$

for all $\varphi \in C(\Delta(\Omega))$.

Lemma 2 *The functional $V : C(\Delta(\Omega)) \to \mathbb{R}$ is a well-defined extension of $V : \Phi_{\mathbb{F}} \to \mathbb{R}$.*

Proof. Take any $\varphi \in C(\Delta(\Omega))$. Since φ is a continuous function defined on a compact set $\Delta(\Omega)$, there exist $p^*, p_* \in \Delta(\Omega)$ such that $\varphi(p^*) \geq \varphi(p) \geq \varphi(p_*)$ for all $p \in \Delta(\Omega)$. Since $\Phi_{\mathbb{F}}$ is a cone including a constant function, $\alpha\mathbf{1} \in \Phi_{\mathbb{F}}$ for all $\alpha \in \mathbb{R}$, where $\mathbf{1} \in C(\Delta(\Omega))$ is the constant function that takes one for all coordinates. Then, for all $\alpha \geq \varphi(p^*)$, $\alpha\mathbf{1} \geq \varphi$. Therefore, $\{V(\varphi_F) | \varphi_F \in \Phi_{\mathbb{F}}, \varphi_F \geq \varphi\} \neq \emptyset$. Moreover, since $\varphi(p) \geq \varphi(p_*)$ for all p, $\varphi \geq \varphi(p_*)\mathbf{1}$. Thus, for every $\varphi_F \geq \varphi$, we have $\varphi_F \geq \varphi(p_*)\mathbf{1}$. By monotonicity of V on $\Phi_{\mathbb{F}}$, $V(\varphi_F) \geq V(\varphi(p_*)\mathbf{1}) = \varphi(p^*)$, that is, $\varphi(p^*)$ is a lower bound for the set. Thus, there exists an infimum, as desired.

To verify that this V is an extension, take any $\varphi_G \in \Phi_{\mathbb{F}}$. For all $\varphi_F \geq \varphi_G$, monotonicity of $V : \Phi_{\mathbb{F}} \to \mathbb{R}$ implies $V(\varphi_F) \geq V(\varphi_G)$. That is, $V(\varphi_G)$ attains the infimum. Therefore, $V(\varphi_G) = \inf\{V(\varphi_F) | \varphi_F \in \Phi_{\mathbb{F}}, \varphi_F \geq \varphi_G\}$. ■

Lemma 3 *The functional $V : C(\Delta(\Omega)) \to \mathbb{R}$ is monotone, normalized, quasi-convex, and continuous.*

Proof. First, note that V is monotone. Moreover, V is also normalized.

Claim 5 *Suppose that $V : C(\Delta(\Omega)) \to \mathbb{R}$ is monotone. Then,*

$$\{\varphi \in C(\Delta(\Omega)) | V(\varphi) < \lambda\} = \{\varphi_F \in \Phi_{\mathbb{F}} | V(\varphi_F) < \lambda\} + C_-(\Delta(\Omega)).$$

Proof. $\supseteq$: Take $\varphi = \varphi_F + \varphi_-$ such that φ_F with $V(\varphi_F) < \lambda$ and $\varphi_- \in C_-(\Delta(\Omega))$. Then, $\varphi \in C(\Delta(\Omega))$. Since $V : C(\Delta(\Omega)) \to \mathbb{R}$ is monotone, $V(\varphi) \leq V(\varphi_F) < \lambda$. Hence, $\varphi \in C(\Delta(\Omega))$ with $V(\varphi) < \lambda$.

$\subseteq$: Take $\varphi \in C(\Delta(\Omega))$ with $V(\varphi) < \lambda$. By the definition of infimum, for any $\varepsilon > 0$, there exists $\varphi_F \in \Phi_{\mathbb{F}}$ such that $\varphi_F \geq \varphi$ and $V(\varphi)+\varepsilon > V(\varphi_F)$. Fix $\varepsilon > 0$ such that $\lambda > V(\varphi)+\varepsilon$. Then, we can find $\varphi_F \in \Phi_{\mathbb{F}}$ such that $\varphi_F \geq \varphi$ and $\lambda > V(\varphi)+\varepsilon > V(\varphi_F)$. Since $\varphi_F \geq \varphi$, we have $\varphi_- = \varphi - \varphi_F \in C_-(\Delta(\Omega))$. Hence, we have $\varphi \in \{\varphi_F \in \Phi_{\mathbb{F}} | V(\varphi_F) < \lambda\} + C_-(\Delta(\Omega))$. ■

Claim 6 *Suppose that $V : C(\Delta(\Omega)) \to \mathbb{R}$ is monotone. If $V : \Phi_{\mathbb{F}} \to \mathbb{R}$ is quasi-convex, $V : C(\Delta(\Omega)) \to \mathbb{R}$ is quasi-convex.*

Proof. We show that $\{\varphi \in C(\Delta(\Omega)) | V(\varphi) < \lambda\}$ is convex for any $\lambda \in \mathbb{R}$. Take $\varphi, \varphi' \in \{\varphi \in C(\Delta(\Omega)) | V(\varphi) < \lambda\}$. Then, by Claim 5, $\varphi = \varphi_F + \varphi_-$ and $\varphi' = \varphi_G + \varphi'_-$ such that $\varphi_F, \varphi_G \in \{\varphi_F \in \Phi_{\mathbb{F}} | V(\varphi_F) < \lambda\}$ and $\varphi_-, \varphi'_- \in C_-(\Delta(\Omega))$. For any $\alpha \in (0,1)$, $V(\alpha\varphi + (1-\alpha)\varphi') = V(\alpha(\varphi_F + \varphi_-) + (1-\alpha)(\varphi_G + \varphi'_-)) \leq V(\alpha\varphi_F + (1-\alpha)\varphi_G) < \lambda$, where the first inequality follows from monotonicity of V, and the second inequality follows from quasi-convexity of $V : \Phi_{\mathbb{F}} \to \mathbb{R}$. Hence, $\{\varphi \in C(\Delta(\Omega)) | V(\varphi) < \lambda\}$ is a convex set for any $\lambda \in \mathbb{R}$. ■

The function $V : C(\Delta(\Omega)) \to \mathbb{R}$ is upper semicontinuous if for any $\lambda > V(\varphi)$ with $\varphi \in C(\Delta(\Omega))$, there exists $\varepsilon > 0$ such that $\lambda > V(\varphi')$ for any $\varphi' \in C(\Delta(\Omega))$ with $\|\varphi - \varphi'\| < \varepsilon$. The function V is upper semi-continuous if and only if $\{\varphi' \in C(\Delta(\Omega)) | V(\varphi') < V(\varphi)\}$ is open. Similarly, the function V is lower semi-continuous if and only if $\{\varphi' \in C(\Delta(\Omega)) \,|\, V(\varphi') > V(\varphi)\}$ is open. The function V is continuous if it is upper and lower semi-continuous.

Claim 7 *The function $V : C(\Delta(\Omega)) \to \mathbb{R}$ is continuous.*

Proof. We show that the function $V : C(\Delta(\Omega)) \to \mathbb{R}$ is upper semi-continuous. This is because Claims 4 and 5 imply that $\{\varphi \in C(\Delta(\Omega)) | V(\varphi) < \lambda\} = \cup_{\varphi_- \in C_-(\Delta(\Omega))}[\{\varphi_F \in \Phi_{\mathbb{F}} | V(\varphi_F) < \lambda\} + \{\varphi_-\}]$ is open. To show this, take $\varphi \in \cup_{\varphi_- \in C_-(\Delta(\Omega))}[\{\varphi_F \in \Phi_{\mathbb{F}} | V(\varphi_F) < \lambda\} + \{\varphi_-\}]$. There exists $\varphi_- \in C_-(\Delta(\Omega))$ such that $\varphi \in \{\varphi_F \in \Phi_{\mathbb{F}} | V(\varphi_F) < \lambda\} + \{\varphi_-\}$. Since $\{\varphi_F \in \Phi_{\mathbb{F}} | V(\varphi_F) < \lambda\} + \{\varphi_-\}$ with $\varphi_- \in C_-(\Delta(\Omega))$ is open, there exists $\varepsilon > 0$ such that any $\varphi' \in C(\Delta(\Omega))$ with $\|\varphi - \varphi'\| < \varepsilon$ satisfies $\varphi' \in \{\varphi_F \in \Phi_{\mathbb{F}} | V(\varphi_F) < \lambda\} + \{\varphi_-\} \subset \cup_{\varphi_- \in C_-(\Delta(\Omega))}[\{\varphi_F \in \Phi_{\mathbb{F}} | V(\varphi_F) < \lambda\} + \{\varphi_-\}]$.

By the symmetric argument, V is lower semi-continuous. ■

■

For all $\pi \in ca_+(\Delta(\Omega))$ and $t \in \mathbb{R}$, define

$$B(\pi, t) = \{\varphi \in C(\Delta(\Omega)) \,|\, \langle \varphi, \pi \rangle \geq t\}, \text{ and}$$
$$W(\pi, t) = \inf_{\varphi \in B(\pi,t)} V(\varphi). \tag{11}$$

Since all constant functions belong to $C(\Delta(\Omega))$, $B(\pi, t) \neq \emptyset$ for all π and t. Thus, $W(\pi, t) < \infty$ for all (π, t), but it is possible that $W(\pi, t) = -\infty$ for some (π, t).

Lemma 4 *For all $\pi \in ca_+(\Delta(\Omega))$, $t \in \mathbb{R}$, and $\alpha > 0$, the following hold:*

(1) $B(\pi, \alpha t) = \alpha B(\pi, t)$.

(2) $B(\alpha\pi, \alpha t) = B(\pi, t)$.

(3) $W(\alpha\pi, \alpha t) = W(\pi, t)$.

Proof. (1) Take any $\varphi \in B(\pi, \alpha t)$. By definition, $\langle \varphi, \pi \rangle \geq \alpha t$, which implies $\langle \varphi/\alpha, \pi \rangle \geq t$. Thus, $\varphi/\alpha \in B(\pi, t)$, or equivalently, $\varphi \in \alpha B(\pi, t)$. The converse is also true.
(2) This part follows from the definition of $B(\pi, t)$.
(3) This follows from part (2). ■

We show that V is rewritten as

$$V(\varphi) = \max_{\pi \in ca_+(\Delta(\Omega))} W(\pi, \langle \varphi, \pi \rangle),$$

which is a counterpart of the "uncertain averse representation"of Cerreia-Vioglio, Maccheroni, Marinacci, and Montrucchio [8] in our setting.

Lemma 5 *For all $\varphi \in C(\Delta(\Omega))$,*

$$V(\varphi) \geq \sup_{\pi \in ca_+(\Delta(\Omega))} W(\pi, \langle \varphi, \pi \rangle).$$

Proof. For every $\pi \in ca_+(\Delta(\Omega))$, we have $\varphi \in B(\pi, \langle \varphi, \pi \rangle)$. By the definition of W, we have $V(\varphi) \geq W(\pi, \langle \varphi, \pi \rangle)$ for any $\pi \in ca_+(\Delta(\Omega))$, and hence $V(\varphi) \geq \sup_{\pi \in ca_+(\Delta(\Omega))} W(\pi, \langle \varphi, \pi \rangle)$. ■

Lemma 6 *For all $\varphi \in C(\Delta(\Omega))$,*

$$V(\varphi) = \max_{\pi \in ca_+(\Delta(\Omega))} W(\pi, \langle \varphi, \pi \rangle).$$

Proof. We modify the proof in Theorem 1 in Cerreia-Vioglio, Maccheroni, Marinacci, and Montrucchio [6] to our setup. We show that there exists $\widetilde{\pi} \in ca_+(\Delta(\Omega))$ such that $V(\varphi) = W(\widetilde{\pi}, \langle \varphi, \widetilde{\pi} \rangle)$. Then, by Lemma 5, we have $V(\varphi) = \max_{\pi \in ca_+(\Delta(\Omega))} W(\pi, \langle \varphi, \pi \rangle)$. Let $SL(\varphi) = \{\varphi' \in C(\Delta(\Omega)) | \, V(\varphi') < V(\varphi)\} \neq \emptyset$. Since V is upper-semi continuous and quasi-convex, $SL(\varphi)$ is convex and open in $C(\Delta(\Omega))$. Since $\varphi \notin SL(\varphi)$, the separation hyperplane theorem ensures that there exists $\widetilde{\pi} \in ca(\Delta(\Omega))$ such that $\langle \varphi, \widetilde{\pi} \rangle > \langle \varphi', \widetilde{\pi} \rangle$ for all $\varphi' \in SL(\varphi)$.

We claim that the separating $\widetilde{\pi} \in ca_+(\Delta(\Omega))$. Fix $\widetilde{\varphi} \in C_+(\Delta(\Omega))$ and $\varphi' \in SL(\varphi)$ arbitrarily. Since V is monotone, $\varphi' \geq \varphi' - n\widetilde{\varphi}$ for all $n \in \mathbb{N}$ implies that $V(\varphi) > V(\varphi') \geq V(\varphi' - n\widetilde{\varphi})$ and hence $\varphi' - n\widetilde{\varphi} \in SL(\varphi)$ for all $n \in \mathbb{N}$. Then, we have that $\langle \varphi, \widetilde{\pi} \rangle > \langle \varphi', \widetilde{\pi} \rangle$ $-n \langle \widetilde{\varphi}, \widetilde{\pi} \rangle$ for all $n \in \mathbb{N}$. Therefore, $\langle \widetilde{\varphi}, \widetilde{\pi} \rangle > \frac{1}{n} (\langle \varphi', \widetilde{\pi} \rangle - \langle \varphi, \widetilde{\pi} \rangle)$ for all $n \in \mathbb{N}$. This implies that $\langle \widetilde{\varphi}, \widetilde{\pi} \rangle \geq 0$ for any $\widetilde{\varphi} \in C_+(\Delta(\Omega))$. Since $\langle \cdot, \widetilde{\pi} \rangle$ is a positive linear functional, Riesz representation theorem implies that there exists a unique $\pi \in ca_+(\Delta(\Omega))$ representing such a positive linear functional. By the uniqueness property, we have $\pi = \widetilde{\pi} \in ca_+(\Delta(\Omega))$.

The property of the separating $\widetilde{\pi} \in ca_+(\Delta(\Omega))$ means that for all φ' with $V(\varphi') < V(\varphi)$, since $\langle \varphi, \widetilde{\pi} \rangle > \langle \varphi', \widetilde{\pi} \rangle$, we have $\varphi' \notin B(\widetilde{\pi}, \langle \varphi, \widetilde{\pi} \rangle)$. By the contraposition, $\varphi' \in B(\widetilde{\pi}, \langle \varphi, \widetilde{\pi} \rangle)$ implies that $V(\varphi') \geq V(\varphi)$. That is,

$$V(\varphi) = \inf_{\varphi' \in B(\widetilde{\pi}, t)} V(\varphi') = W(\widetilde{\pi}, \langle \varphi, \widetilde{\pi} \rangle).$$

■

By Lemmas 4 and 6, we conclude that

$$V(\varphi) = \max_{\pi \in \Delta(\Delta(\Omega))} W(\pi, \langle \varphi, \pi \rangle).$$

In particular, $\succsim$ is represented by

$$U(F) = V(\varphi_F) = \max_{\pi \in \Delta(\Delta(\Omega))} W(\pi, \langle \varphi_F, \pi \rangle) = \max_{\pi \in \Delta(\Delta(\Omega))} W(\pi, b_F^u(\pi)).$$

We show several properties of W.

Lemma 7 *(1) For any $\pi \in \Delta(\Delta(\Omega))$, $W(\pi, t)$ is nondecreasing in t.*

(2) $W(\pi, t)$ is quasi-concave in $(\pi, t) \in \Delta(\Delta(\Omega)) \times \mathbb{R}$.

(3) $W(\pi, t)$ is upper semi-continuous in $(\pi, t) \in \Delta(\Delta(\Omega)) \times \mathbb{R}$.

Proof. (1) Take t and t' with $t > t'$. Since $B(\pi, t) \subsetneq B(\pi, t')$, we have $W(\pi, t) \geq W(\pi, t')$.

(2) The proof follows from Cerreia-Vioglio, Maccheroni, Marinacci, and Montrucchio [6, Lemma 4]. Take any (π_i, t_i) for $i = 1, 2$ and $\alpha \in [0, 1]$. Let $\pi' = \alpha\pi_1 + (1 - \alpha)\pi_2$ and $t' = \alpha t_1 + (1 - \alpha)t_2$. Then,

$$B(\pi', t') \subset B(\pi_1, t_1) \cup B(\pi_2, t_2),$$

which implies

$$W(\pi', t') \geq \inf_{\varphi \in B(\pi_1, t_1) \cup B(\pi_2, t_2)} V(\varphi) = \min[W(\pi_1, t_1), W(\pi_2, t_2)],$$

which means that W is quasi-concave.

(3) The proof follows from Cerreia-Vioglio, Maccheroni, Marinacci, and Montrucchio [6, Lemma 5]. Take any $\overline{\pi} \in \Delta(\Delta(\Omega))$ and $\alpha, \overline{t} \in \mathbb{R}$ such that $W(\overline{\pi}, \overline{t}) < \alpha$. There exists $\varphi_0 \in C(\Delta(\Omega))$ such that $\langle \varphi_0, \overline{\pi} \rangle \geq \overline{t}$ and $V(\varphi_0) < \alpha$. The sequence $\varphi^n = \varphi_0 + \frac{1}{n}\mathbf{1}$ converges to φ_0 as $n \to \infty$. Since V is upper semi-continuous, there exists $\overline{n}$ such that $V(\varphi^{\overline{n}}) < \alpha$. Moreover,

$$\langle \varphi^{\overline{n}}, \overline{\pi} \rangle = \langle \varphi_0, \overline{\pi} \rangle + \frac{1}{\overline{n}} \langle \mathbf{1}, \overline{\pi} \rangle \geq \overline{t} + \frac{1}{\overline{n}}.$$

Note that the set

$$O = \{\pi \in \Delta(\Delta(\Omega)) \,|\, \langle \varphi^{\overline{n}}, \pi \rangle > \langle \varphi^{\overline{n}}, \overline{\pi} \rangle - \frac{1}{2\overline{n}}\}$$

is open in the topology induced by the weak* topology. It is easy to see that $O \times (-\infty, \overline{t} + \frac{1}{2\overline{n}})$ is an open neighborhood of $(\overline{\pi}, \overline{t})$. Moreover, for all $(\pi, t) \in O \times (-\infty, \overline{t} + \frac{1}{2\overline{n}})$, we have

$$\langle \varphi^{\overline{n}}, \pi \rangle > \langle \varphi^{\overline{n}}, \overline{\pi} \rangle - \frac{1}{2\overline{n}} \geq \overline{t} + \frac{1}{\overline{n}} - \frac{1}{2\overline{n}} = \overline{t} + \frac{1}{2\overline{n}} > t.$$

Hence, $W(\pi, t) \leq V(\varphi^{\overline{n}}) < \alpha$, and $W(\pi, t)$ is upper semi-continuous. ■

Lemma 8 *W is linearly continuous.*

Proof. As shown in Lemma 3, V is continuous on $C(\Delta(\Omega))$. Moreover, V is written as

$$V(\varphi) = \max_{\pi \in \Delta(\Delta(\Omega))} W(\pi, \langle \varphi, \pi \rangle),$$

which implies that W is linearly continuous. ■

Lemma 9 *For all $\pi \in \Delta(\Delta(\Omega))$ and $t \in \mathbb{R}$,*

$$W(\pi, t) = \inf_{\varphi_F \in B(\pi, t)} V(\varphi_F) = \inf_{\{F \,|\, b^u_F(\pi) \geq t\}} u(x_F).$$

Proof. It is enough to show that

$$\inf_{\varphi \in B(\pi, t)} V(\varphi) = \inf_{\varphi_F \in B(\pi, t)} V(\varphi_F).$$

For all $\varphi \in C(\Delta(\Omega))$, define $D(\varphi) = \{\varphi_F \in \Phi_{\mathbb{F}} : \varphi_F \geq \varphi\}$. Since $B(\pi, t) \cap D(\varphi) \subset B(\pi, t)$, $\inf_{\varphi \in B(\pi, t)} V(\varphi) \leq \inf_{\varphi_F \in B(\pi, t)} V(\varphi_F)$. Hence, it is enough to show the converse, that is, $\inf_{\varphi \in B(\pi, t)} V(\varphi) \geq \inf_{\varphi_F \in B(\pi, t)} V(\varphi_F)$.

Take any $\varepsilon > 0$. By definition of infimum, there exists $\varphi^\varepsilon \in B(\pi, t)$ such that $V(\varphi^\varepsilon) < \inf_{\varphi \in B(\pi, t)} V(\varphi) + \varepsilon$. By definition of V, $\inf_{\varphi_F \in D(\varphi^\varepsilon)} V(\varphi_F) < \inf_{\varphi \in B(\pi, t)} V(\varphi) + \varepsilon$. Again, by definition of infimum, there exists $\varphi_F^\varepsilon \in D(\varphi^\varepsilon)$ such that $V(\varphi_F^\varepsilon) < \inf_{\varphi \in B(\pi,t)} V(\varphi) + \varepsilon$. Moreover, since $\varphi_F^\varepsilon \geq \varphi^\varepsilon$ and $\varphi^\varepsilon \in B(\pi, t)$, we have $\langle \varphi_F^\varepsilon, \pi\rangle \geq \langle \varphi^\varepsilon, \pi \rangle \geq t$, that is, $\varphi_F^\varepsilon \in B(\pi, t)$. Consequently, for all $\varepsilon > 0$, we can find some $\varphi_F^\varepsilon \in B(\pi, t)$ such that

$$V(\varphi_F^\varepsilon) < \inf_{\varphi \in B(\pi,t)} V(\varphi) + \varepsilon.$$

By definition of infimum,

$$\inf_{\varphi_F \in B(\pi,t)} V(\varphi_F) < \inf_{\varphi \in B(\pi,t)} V(\varphi) + \varepsilon.$$

Therefore, we have the desired result as $\varepsilon \to 0$. ■

Lemma 10 $W(\delta_{\overline{p}}, t) = t$.

Proof. Take a lottery x whose value is $u(x) = t$. The representation implies

$$t = u(x) = V(\varphi_{\{x\}}) = \max_{\pi \in \Delta(\Delta(\Omega))} W(\pi, b^u_{\{x\}}(\pi)) = \max_{\pi \in \Delta(\Delta(\Omega))} W(\pi, t).$$

Thus, $t \geq W(\pi, t)$ for all $\pi \in \Delta(\Delta(\Omega))$.

It is enough to show that $W(\delta_{\overline{p}}, t) \geq t$. For all menus F,

$$\langle \varphi_F, \delta_{\overline{p}} \rangle = \max_{f \in F} \sum_\omega u(f(\omega))\overline{p}(\omega) = \sum_\omega u(f^F(\omega))\overline{p}(\omega),$$

where $f^F \in F$ is a maximizer. By monotonicity of V,

$$V(\varphi_F) \geq V(\varphi_{\{f^F\}}) = U(\{f^F\}) = \sum_\omega u(f^F(\omega))\overline{p}(\omega) = \langle \varphi_F, \delta_{\overline{p}} \rangle.$$

Thus, by Lemma 9,

$$W(\delta_{\overline{p}}, t) = \inf_{\varphi \in B(\delta_{\overline{p}},t)} V(\varphi) = \inf_{\varphi_F \in B(\delta_{\overline{p}},t)} V(\varphi_F) \geq \inf_{\varphi_F \in B(\delta_{\overline{p}},t)} \langle \varphi_F, \delta_{\overline{p}} \rangle \geq t.$$

as desired. ■

Lemma 11 *If* $\pi \trianglerighteq \rho$, $W(\pi, t) \leq W(\rho, t)$ *for all* t.

Proof. If $\pi \trianglerighteq \rho$, $\langle \varphi_F, \pi \rangle \geq \langle \varphi_F, \rho \rangle$ for all menus F. Thus, $B(\rho, t) \subset B(\pi, t)$, which implies, together with Lemma 9,

$$W(\pi, t) = \inf_{\varphi_F \in B(\pi,t)} V(\varphi_F) \leq \inf_{\varphi_F \in B(\rho,t)} V(\varphi_F) = W(\rho, t).$$

■

Define

$$\Pi = \{\pi \in \Delta(\Delta(\Omega)) \,|\, W(\pi, t) > -\infty \text{ for some } t\}. \tag{12}$$

By Lemma 10, $\delta_{\overline{p}} \in \Pi$. In particular, $\Pi \neq \emptyset$. Since any $\pi \notin \Pi$ never achieves the maximum of W, the representation U is rewritten as

$$U(F) = V(\varphi_F) = \max_{\pi \in \Pi} W(\pi, \langle \varphi_F, \pi \rangle).$$

Finally, we show the Bayesian plausibility condition of Π. For all $\pi \in \Pi$, let

$$p^\pi = \int_{\Delta(\Omega)} p \, \mathrm{d}\pi(p) \in \Delta(\Omega).$$

Lemma 12 $\Pi \subset \Pi(\overline{p})$.

Proof. We show that for all $\pi \in \Pi$, $p^\pi = \overline{p}$. Seeking a contradiction, suppose that there exists $\pi^* \in \Pi$ such that $p^{\pi^*} \neq \overline{p}$. There exist ω and ω' such that $p^{\pi^*}_\omega > \overline{p}_\omega$ and $p^{\pi^*}_{\omega'} < \overline{p}_{\omega'}$.

By definition of Π, there exists some t^* such that $W(\pi^*, t^*) > -\infty$. Take any $a < W(\pi^*, t^*)$. Let f be a constant act satisfying $u(f(\omega)) = a$ for all ω. Let $\mathbf{a}$ denote the vector in $\mathbb{R}^\Omega$ which takes a value of a for all coordinates. Consider a vector

$$\widetilde{\mathbf{a}} = (a, \cdots, a, \widetilde{a}_\omega, a, \cdots, a, \widetilde{a}_{\omega'}, a, \cdots, a) \in \mathbb{R}^\Omega$$

which satisfies $\widetilde{\mathbf{a}} \cdot \overline{p} = a$. Since $\mathbf{a} \cdot \overline{p} = a$, $\widetilde{\mathbf{a}}$ can be regarded as a utility act which is indifferent to a constant utility act $\mathbf{a}$ under the subjective expected utility with $\overline{p}$. From $\widetilde{\mathbf{a}} \cdot \overline{p} = a$, we have

$$\overline{p}_\omega \widetilde{a}_\omega + \overline{p}_{\omega'} \widetilde{a}_{\omega'} = \overline{p}_\omega a + \overline{p}_{\omega'} a,$$

or

$$\widetilde{a}_{\omega'} = A - \frac{\overline{p}_\omega}{\overline{p}_{\omega'}} \widetilde{a}_\omega, \text{ where } A = \frac{\overline{p}_\omega a + \overline{p}_{\omega'} a}{\overline{p}_{\omega'}}. \tag{13}$$

From (13),

$$\begin{aligned}
\widetilde{\mathbf{a}} \cdot p^{\pi^*} &= p^{\pi^*}_\omega \widetilde{a}_\omega + p^{\pi^*}_{\omega'} \widetilde{a}_{\omega'} + (1 - p^{\pi^*}_\omega + p^{\pi^*}_{\omega'}) a \\
&= (p^{\pi^*}_\omega - \overline{p}_\omega) \widetilde{a}_\omega + (p^{\pi^*}_{\omega'} - \overline{p}_{\omega'}) \widetilde{a}_{\omega'} + \overline{p}_\omega \widetilde{a}_\omega + \overline{p}_{\omega'} \widetilde{a}_{\omega'} + (1 - p^{\pi^*}_\omega + p^{\pi^*}_{\omega'}) a \\
&= (p^{\pi^*}_\omega - \overline{p}_\omega) \widetilde{a}_\omega + (p^{\pi^*}_{\omega'} - \overline{p}_{\omega'}) \widetilde{a}_{\omega'} + \overline{p}_\omega a + \overline{p}_{\omega'} a + (1 - p^{\pi^*}_\omega + p^{\pi^*}_{\omega'}) a \\
&= \left\{ (p^{\pi^*}_\omega - \overline{p}_\omega) - \frac{\overline{p}_\omega}{\overline{p}_{\omega'}} (p^{\pi^*}_{\omega'} - \overline{p}_{\omega'}) \right\} \widetilde{a}_\omega + B \\
&= \frac{\overline{p}_{\omega'} p^{\pi^*}_\omega - \overline{p}_\omega p^{\pi^*}_{\omega'}}{\overline{p}_{\omega'}} \widetilde{a}_\omega + B,
\end{aligned} \tag{14}$$

where

$$B = (p^{\pi^*}_{\omega'} - \overline{p}_{\omega'}) A + \overline{p}_\omega a + \overline{p}_{\omega'} a + (1 - p^{\pi^*}_\omega + p^{\pi^*}_{\omega'}) a.$$

Since $p^{\pi^*}_\omega > \overline{p}_\omega$ and $p^{\pi^*}_{\omega'} < \overline{p}_{\omega'}$, the multiplier of $\widetilde{a}_\omega$ in (14) is positive. Since (14) is a positive linear function with respect to $\widetilde{a}_\omega$, $\widetilde{\mathbf{a}} \cdot p^{\pi^*}$ varies across all the real numbers. By choosing $\widetilde{a}_\omega$ appropriately, we can set $\widetilde{\mathbf{a}} \cdot p^{\pi^*} = t^*$.

Since $u(X) = \mathbb{R}$, we can find some $\widetilde{f} \in \mathcal{F}$ satisfying $u(\widetilde{f}) = \widetilde{\mathbf{a}}$. By construction, $U(\{\widetilde{f}\}) = \widetilde{\mathbf{a}} \cdot \overline{p} = a$. However, by assumption,

$$U(\{\widetilde{f}\}) = a < W(\pi^*, t^*) = W(\pi^*, \widetilde{\mathbf{a}} \cdot p^{\pi^*}) \leq \max_{\pi \in \Pi} W(\pi, \widetilde{\mathbf{a}} \cdot p^{\pi}) = \max_{\pi \in \Pi} W(\pi, \langle \varphi_{\{\widetilde{f}\}}, \pi \rangle),$$

which contradicts to the representation. ■

By Lemma 12,

$$U(F) = V(\varphi_F) = \max_{\pi \in \Pi(\overline{p})} W(\pi, \langle \varphi_F, \pi \rangle)$$

is a Costly Subjective Learning Representation.

B.2 Necessity

Now, we show that the axioms are satisfied. Let $\succsim$ be the preference U represents. It is obvious that $\succsim$ is complete and transitive. Since $u(X) = \mathbb{R}$, $\succsim$ satisfies Two-Sided Unboundedness.

Mixture Continuity

Take any $F, G \in \mathbb{F}$ and $\alpha \in [0, 1]$. From the representation,

$$\begin{aligned} U(\alpha F + (1-\alpha)G) &= \max_{\pi \in \Pi} W(\pi, b^u_{\alpha F + (1-\alpha)G}(\pi)) \\ &= \max_{\pi \in \Pi} W(\pi, \langle \alpha \varphi_F + (1-\alpha)\varphi_G, \pi \rangle). \end{aligned}$$

Since W is linearly continuous, $U(\alpha F + (1-\alpha)G)$ is continuous in α. Hence, $U(\alpha^n F + (1-\alpha^n)G) \to U(\alpha F + (1-\alpha)G)$ as $\alpha^n \to \alpha$, which implies Mixture Continuity of $\succsim$.

Preference for Flexibility

Take any F and G with $G \subset F$. We have $b^u_F(\pi) \geq b^u_G(\pi)$ for all π. Since $W(\pi, t)$ is non-decreasing in t,

$$U(F) = \max_{\pi \in \Pi} W(\pi, b^u_F(\pi)) \geq \max_{\pi \in \Pi} W(\pi, b^u_G(\pi)) = U(G),$$

which implies that $\succsim$ satisfies Preference for Flexibility.

Dominance

Take any F and g. Assume that there exists $f \in F$ such that $\{f(\omega)\} \succsim \{g(\omega)\}$ for all ω. Since $b^u_F(\pi) = b^u_{F \cup \{g\}}(\pi)$ for all π,

$$U(F) = \max_{\pi \in \Pi} W(\pi, b^u_F(\pi)) = \max_{\pi \in \Pi} W(\pi, b^u_{F \cup \{g\}}(\pi)) = U(F \cup \{g\}).$$

Thus, Dominance holds.

Singleton Independence

For any $f \in \mathcal{F}$ and $\pi \in \Pi$, we have

$$b^u_{\{f\}}(\pi) = \sum_{\Omega} u(f(\omega))p^\pi(\omega) = \sum_{\Omega} u(f(\omega))\overline{p}(\omega).$$

Since $\pi \trianglerighteq \delta_{\overline{p}}$ for all $\pi \in \Pi$, we have $t = W(\delta_{\overline{p}}, t) = \max_{\pi\in\Pi} W(\pi, t)$, which implies

$$U(\{f\}) = \max_{\pi\in\Pi} W(\pi, b^u_{\{f\}}(\pi)) = \max_{\pi\in\Pi} W(\pi, \sum_{\Omega} u(f(\omega))\overline{p}(\omega)) = \sum_{\Omega} u(f(\omega))\overline{p}(\omega).$$

Since $U(\{f\})$ is a subjective expected utility function, it satisfies Singleton Independence.

Aversion to Contingent Planning

The proof follows from Cerreia-Vioglio, Maccheroni, Marinacci, and Montrucchio [8, Lemma 47]. It suffices to show that $V(\varphi) = \sup_{\pi\in\Pi} W(\pi, \langle\varphi, \pi\rangle)$, which is defined on $\Delta(\Delta(\Omega))$, is quasi-convex. In fact, if this is the case, for any F, G, and $\alpha \in (0, 1)$,

$$\begin{aligned} U(\alpha F + (1-\alpha)G) &= \sup_{\pi\in\Pi} W(\pi, \langle\varphi_{\alpha F+(1-\alpha)G}, \pi\rangle) \\ &= \sup_{\pi\in\Pi} W(\pi, \langle\alpha\varphi_F + (1-\alpha)\varphi_G, \pi\rangle) \\ &\leq \max[\sup_{\pi\in\Pi} W(\pi, \langle\varphi_F, \pi\rangle), \sup_{\pi\in\Pi} W(\pi, \langle\varphi_F, \pi\rangle)] \\ &= \max[U(F), U(G)]. \end{aligned}$$

Now take any $t \in \mathbb{R}$. We want to show that $\{\varphi \,|\, V(\varphi) \leq t\}$ is convex. Define

$$L = \bigcap_{(\pi,t')\in\Pi\times\mathbb{R}\,:\,\{\varphi \,|\, \langle\varphi,\pi\rangle<t'\}\supset\{\varphi \,|\, V(\varphi)\leq t\}} \{\varphi \,|\, \langle\varphi, \pi\rangle < t'\}.$$

Note that L is a convex set because it is the intersection of a family of open half spaces. Moreover, by definition, $\{\varphi \,|\, V(\varphi) \leq t\} \subset L$. We will show the converse, whereby establishing $L = \{\varphi \,|\, V(\varphi) \leq t\}$, and hence, $\{\varphi \,|\, V(\varphi) \leq t\}$ is convex, as desired.

Take any $\overline{\varphi} \notin \{\varphi \,|\, V(\varphi) \leq t\}$. Then, $t < V(\varphi) = \max_{\pi\in\Pi} W(\pi, \langle\overline{\varphi}, \pi\rangle)$. There exists $\overline{\pi} \in \Pi$ such that $t < W(\overline{\pi}, \langle\overline{\varphi}, \overline{\pi}\rangle)$. For any φ with $\langle\varphi, \overline{\pi}\rangle \geq \langle\overline{\varphi}, \overline{\pi}\rangle$, we have

$$t < W(\overline{\pi}, \langle\overline{\varphi}, \overline{\pi}\rangle) \leq W(\overline{\pi}, \langle\varphi, \overline{\pi}\rangle) \leq V(\varphi).$$

That is,

$$\{\varphi \,|\, \langle\varphi, \overline{\pi}\rangle \geq \langle\overline{\varphi}, \overline{\pi}\rangle\} \subset \{\varphi \,|\, V(\varphi) > t\},$$

or equivalently,

$$\{\varphi \,|\, \langle\varphi, \overline{\pi}\rangle < \langle\overline{\varphi}, \overline{\pi}\rangle\} \supset \{\varphi \,|\, V(\varphi) \leq t\}.$$

Since $\overline{\varphi} \notin \{\varphi \,|\, \langle\varphi, \overline{\pi}\rangle < \langle\overline{\varphi}, \overline{\pi}\rangle\}$, by definition of L, $\overline{\varphi} \notin L$. Therefore, $L \subset \{\varphi \,|\, V(\varphi) \leq t\}$, as desired.

C Proof of Theorem 3

$[(a) \implies (b)]$ It is easy to see that if $\succsim_1$ is more averse to commitment than $\succsim_2$, then for all $f, g \in \mathcal{F}$, $\{f\} \succsim_1 \{g\}$ if and only if $\{f\} \succsim_2 \{g\}$, that is, the two preferences are identical on singleton menus. Thus, by Anscombe and Aumann [2], there exist $\alpha > 0$ and $\beta \in \mathbb{R}$ such that $u_2 = \alpha u_1 + \beta$ and that $\overline{p}_1 = \overline{p}_2$. Without loss of generality, we assume that $u_1 = u_2 = u$. Since $\succsim_1$ is more averse to commitment than $\succsim_2$, for any $F \in \mathbb{F}$ and $f \in \mathcal{F}$, $F \sim_2 \{f\}$ implies $F \succsim_1 \{f\}$. Let $x_F^1, x_F^2 \in X$ be $\{x_F^1\} \sim_1 F$ and $\{x_F^2\} \sim_2 F$. Hence, $\{x_F^1\} \sim_1 F \succsim_1 \{x_F^2\}$, which implies $u(x_F^1) \geq u(x_F^2)$ for any $F \in \mathbb{F}$. Since $W_i(\pi, t) = \inf_{\{F \,|\, b_F^u(\pi) \geq t\}} u(x_F^i)$ for $i = 1, 2$,

$$W_1(\pi, t) = \inf_{\{F \,|\, b_F^u(\pi) \geq t\}} u(x_F^1) \geq \inf_{\{F \,|\, b_F^u(\pi) \geq t\}} u(x_F^2) = W_2(\pi, t)$$

for any (π, t).

$[(b) \implies (a)]$ Without loss of generality, assume that $u_1 = u_2 = u$. For any $F \in \mathbb{F}$ and $f \in \mathcal{F}$, $F \succsim_2 \{f\}$ implies that

$$\max_{\pi \in \Pi(\overline{p})} W_2(\pi, b_F^u(\pi)) \geq \max_{\pi \in \Pi(\overline{p})} W_2(\pi, b_{\{f\}}^u(\pi)) = \sum_{\omega} u(f(\omega))\overline{p}(\omega).$$

Since $W_1(\pi, t) \geq W_2(\pi, t)$ for all $(\pi, t) \in \Pi(\overline{p}) \times \mathbb{R}$,

$$\max_{\pi \in \Pi(\overline{p})} W_1(\pi, b_F^u(\pi)) \geq \max_{\pi \in \Pi(\overline{p})} W_2(\pi, b_F^u(\pi)) \geq \sum_{\omega} u(f(\omega))\overline{p}(\omega) = \max_{\pi \in \Pi(\overline{p})} W_1(\pi, b_{\{f\}}^u(\pi)).$$

Hence, we have that $F \succsim_1 \{f\}$.

D Proof of Corollary 3

We define $V : \Phi_{\mathbb{F}} \to \mathbb{R}$ as in the proof of Theorem 1. By Lemma 1, it is normalized and monotone.

Lemma 13 *Under Increasing Desire for Commitment and Singleton Independence,* $V : \Phi_{\mathbb{F}} \to \mathbb{R}$ *is convex.*

Proof. Let $\{x_F\}$ be a lottery equivalent of F, that is, $\{x_F\} \sim F$. The existence of a lottery equivalent is guaranteed under Order, Continuity, Monotonicity, and Dominance. By Singleton Independence, $V(\alpha\{x_F\} + (1-\alpha)\{x_G\}) = \alpha V(\{x_F\}) + (1-\alpha)V(\{x_G\}) = \alpha V(F) + (1-\alpha)V(G)$. Hence, $\alpha V(F) + (1-\alpha)V(G) \geq V(\alpha F + (1-\alpha)G)$. ■

We use the same extension of V to $C(\Delta(\Omega))$ as in the proof of Theorem 1. By Lemmas 2 and 3, V on $C(\Delta(\Omega))$ is well-defined, monotone, and normalized.

Lemma 14 *Suppose that* $V : C(\Delta(\Omega)) \to \mathbb{R}$ *is monotone. Then,*

$$\{(\varphi, \widetilde{\lambda}) \in C(\Delta(\Omega)) \times \mathbb{R} \Big| V(\varphi) < \widetilde{\lambda}\} = \{(\varphi_F, \lambda) \in \Phi_{\mathbb{F}} \times \mathbb{R} | V(\varphi_F) < \lambda\} + C_-(\Delta(\Omega)) \times \{0\}.$$

Proof. $\supseteq$: Take $\varphi = \varphi_F + \varphi$ and $\lambda \in \mathbb{R}$ such that φ_F with $V(\varphi_F) < \lambda$ and $\varphi_- \in C_-(\Delta(\Omega))$. Then, $\varphi \in C(\Delta(\Omega))$. Since $V : \Phi_{\mathbb{F}} \to \mathbb{R}$ is monotone, $V(\varphi) \leq V(\varphi_F) < \lambda$. Hence, $\varphi \in C(\Delta(\Omega))$ with $V(\varphi) < \lambda$.

$\subseteq$: Take $(\varphi, \widetilde{\lambda}) \in C(\Delta(\Omega)) \times \mathbb{R}$ with $V(\varphi) < \widetilde{\lambda}$. By the definition of infimum, for any $\varepsilon > 0$, there exists $\varphi_F \in \Phi_{\mathbb{F}}$ such that $\varphi_F \geq \varphi$ and $V(\varphi) + \varepsilon > V(\varphi_F)$. Fix $\varepsilon > 0$ such that $\widetilde{\lambda} > V(\varphi) + \varepsilon$. Then, we can find $\varphi_F \in \Phi_{\mathbb{F}}$ such that $\varphi_F \geq \varphi$ and $\widetilde{\lambda} > V(\varphi) + \varepsilon > V(\varphi_F)$. Since $\varphi_F \geq \varphi$, we have $\varphi_- = \varphi - \varphi_F \in C_-(\Delta(\Omega))$. Hence, we have $(\varphi, \widetilde{\lambda}) \in \{(\varphi_F, \lambda) \in \Phi_{\mathbb{F}} \times \mathbb{R} | V(\varphi_F) < \lambda\} + C_-(\Delta(\Omega)) \times \{0\}$. ■

Lemma 15 *If* $V : \Phi_{\mathbb{F}} \to \mathbb{R}$ *is monotone and convex,* $V : C(\Delta(\Omega)) \to \mathbb{R}$ *is convex.*

Proof. We show that $\{(\varphi, \lambda) \in C(\Delta(\Omega)) \times \mathbb{R} | V(\varphi) < \lambda\}$ is a convex set. Take $(\varphi, \lambda), (\varphi', \lambda') \in \{(\varphi, \lambda) \in C(\Delta(\Omega)) \times \mathbb{R} | V(\varphi) < \lambda\}$, where $\varphi = \varphi_F + \varphi_-$ and $\varphi' = \varphi_G + \varphi'_-$ such that $(\varphi_F, \lambda), (\varphi_G, \lambda') \in \{(\varphi_F, \lambda) \in \Phi_{\mathbb{F}} \times \mathbb{R} | V(\varphi_F) < \lambda\}$ and $\varphi_-, \varphi'_- \in C_-(\Delta(\Omega))$. For any $\alpha \in (0,1)$, $V(\alpha\varphi + (1-\alpha)\varphi') = V(\alpha(\varphi_F + \varphi_-) + (1-\alpha)(\varphi_G + \varphi'_-)) \leq V(\alpha\varphi_F + (1-\alpha)\varphi_G) \leq \alpha V(\varphi_F) + (1-\alpha)V(\varphi_G) < \alpha\lambda + (1-\alpha)\lambda'$, where the first inequality follows from monotonicity of V, and the second inequality follows from convexity of $V : \Phi_{\mathbb{F}} \to \mathbb{R}$. Hence, $\{(\varphi, \lambda) \in C(\Delta(\Omega)) \times \mathbb{R} | V(\varphi) < \lambda\}$ is a convex set. ■

Lemma 16 $V : C(\Delta(\Omega)) \to \mathbb{R}$ *is normalized, monotone, convex, and continuous.*

Proof. The first two properties are already shown. By Lemma 15, V is convex.

Finally, we show that V is continuous. Since $V : C(\Delta(\Omega)) \to \mathbb{R}$ is convex, by Barbu and Precupanu [3, Theorem 2.14], it suffices to show that there exists a neighborhood B at the origin $\mathbf{0} \in C(\Delta(\Omega))$ such that V is bounded from above on B, that is, there exists $k \in \mathbb{R}$ such that $V(\varphi) \leq k$ for all $\varphi \in B$. Take any k-neighborhood at $\mathbf{0}$, denoted by $B_k(\mathbf{0})$. For all $\varphi \in B_k(\mathbf{0})$, we have $\varphi(p) \leq \sup_{p \in \Delta} |\varphi(p)| \leq k$. That is, $\varphi(p) \leq k\mathbf{1}$. Since V is normalized and monotone, $V(\varphi) \leq V(k\mathbf{1}) = k$, as desired. ■

Since $V : C(\Delta(\Omega)) \to \mathbb{R}$ is monotone, normalized, convex, continuous, and $V(C(\Delta(\Omega))) = \mathbb{R}$, by a minor modification of Strzalecki [30, Theorem 3], $V : C(\Delta(\Omega)) \to \mathbb{R}$ is niveloid: for all $\varphi, \varphi' \in C(\Delta(\Omega))$,

$$V(\varphi) - V(\varphi') \leq \sup(\varphi - \varphi').$$

Since $V : C(\Delta(\Omega)) \to \mathbb{R}$ is monotone, normalized, and convex niveloid, de Oliveira, Denti, Mihm, and Ozbek [15, Theorem 1] implies that preference $\succsim$ is represented by an additive cost representation.

E Proof of Proposition 1

(i) By the definition of v^L, we have $b^u_{F+v}(\pi^H) \geq b^u_{F+v}(\pi^L) > 0$ for any $v > v^L$. Thus $W_{F,\pi^L}(v) = \beta(\pi^L)b^u_{F+v}(\pi)$ and $W_{F,\pi^H}(v) = \beta(\pi^H)b^u_{F+v}(\pi^H)$. The assumption implies

$$\frac{\beta(\pi^L)}{\beta(\pi^H)} \geq \frac{b^u_{F+v}(\pi^H)}{b^u_{F+v}(\pi^L)} = \frac{b^u_F(\pi^H)+v}{b^u_F(\pi^L)+v}.$$

Note that the function of the form $f(x) = \frac{a+x}{b+x}$, where $x > b$ and $a \geq b$, is nonincreasing in x. Moreover, $b^u_F(\pi^H) \geq b^u_F(\pi^L)$. Hence,

$$\frac{\beta(\pi^L)}{\beta(\pi^H)} \geq \frac{b^u_F(\pi^H)+v}{b^u_F(\pi^L)+v} \geq \frac{b^u_F(\pi^H)+v'}{b^u_F(\pi^L)+v'} = \frac{b^u_{F+v'}(\pi^H)}{b^u_{F+v'}(\pi^L)},$$

that is, $\beta(\pi^L)b^u_{F+v'}(\pi^L) \geq \beta(\pi^H)b^u_{F+v'}(\pi^H)$, as desired. For the strict inequality, repeat the same argument.

(ii) By the definition of v^H and v^L, we have $b^u_{F+v}(\pi^H) \geq 0 \geq b^u_{F+v}(\pi^L)$ for any $v \in [v^H, v^L]$. Thus $W_{F,\pi^H}(v) = \beta(\pi^H)b^u_{F+v}(\pi^H) \geq 0$ and $W_{F,\pi^L}(v) = \gamma(\pi^L)b^u_{F+v}(\pi^L) \leq 0$.

(iii) By the definition of v^H, we have $b^u_{F+v}(\pi^L) \leq b^u_{F+v}(\pi^H) < 0$ for any $v < v^H$. Thus $W_{F,\pi^L}(v) = \gamma(\pi^L)b^u_{F+v}(\pi)$ and $W_{F,\pi^H}(v) = \gamma(\pi^H)b^u_{F+v}(\pi^H)$. We show the contraposition. The presumption implies

$$\frac{\gamma(\pi^H)}{\gamma(\pi^L)} \leq \frac{b^u_{F+v}(\pi^L)}{b^u_{F+v}(\pi^H)} = \frac{b^u_F(\pi^L)+v}{b^u_F(\pi^H)+v}.$$

Note that the function of the form $f(x) = \frac{b+x}{a+x}$, where $x < a$ and $a \geq b$, is nondecreasing in x. Since $b^u_F(\pi^H) \geq b^u_F(\pi^L)$,

$$\frac{\gamma(\pi^H)}{\gamma(\pi^L)} \leq \frac{b^u_F(\pi^L)+v}{b^u_F(\pi^H)+v} \leq \frac{b^u_F(\pi^L)+v'}{b^u_F(\pi^H)+v'} = \frac{b^u_{F+v'}(\pi^L)}{b^u_{F+v'}(\pi^H)},$$

that is, $\gamma(\pi^H)b^u_{F+v'}(\pi^H) \geq \gamma(\pi^L)b^u_{F+v'}(\pi^L)$, as desired. For the strict inequality, repeat the same argument.

F Proof of Theorem 4

We define $V : \Phi_{\mathbb{F}} \to \mathbb{R}$ as in the proof of Theorem 1. As shown in Lemma 1, V is monotone, normalized, quasi-convex, and continuous. Moreover, V satisfies the following property:

Lemma 17 *$V : \Phi_{\mathbb{F}} \to \mathbb{R}$ is positively homogeneous.*

Proof. We show positive homogeneity of V. For $\alpha \in [0,1]$,

$$\begin{aligned} V(\alpha\varphi_F) &= V(\alpha\varphi_F + (1-\alpha)0) = V(\varphi_{\alpha F+(1-\alpha)\{x_0\}}) \\ &= U(\alpha F + (1-\alpha)\{x_0\}) = U(\alpha\{x_F\} + (1-\alpha)\{x_0\}) \\ &= \alpha U(\{x_F\}) + (1-\alpha)0 = \alpha V(x_F). \end{aligned}$$

The second equality follows from linearity of φ. The forth equality follows from Reference-Point Independence.

For $\alpha \in (1,\infty)$, denote $\varphi_G = \alpha\varphi_F \in \Phi_{\mathbb{F}}$. By the above property,

$$V(\varphi_F) = V(\frac{1}{\alpha}\varphi_G) = \frac{1}{\alpha}V(\varphi_G) = \frac{1}{\alpha}V(\alpha\varphi_F),$$

as desired. ■

As in (10), V is extended to $C(\Delta(\Omega))$. As shown in Lemma 3, $V : C(\Delta(\Omega)) \to \mathbb{R}$ is monotone, normalized, quasi-convex, and continuous.

Lemma 18 $V : C(\Delta(\Omega)) \to \mathbb{R}$ *is positively homogeneous.*

Proof. We show that V satisfies positive homogeneity. For every $\varphi \in C(\Delta(\Omega))$ and $\alpha > 0$, note that

$$\{\varphi_F \in \Phi_{\mathbb{F}} \,|\, \varphi_F \geq \varphi\} = \{\alpha\varphi_F \in \Phi_{\mathbb{F}} \,|\, \alpha\varphi_F \geq \varphi \text{ for some } \varphi_F \in \Phi_{\mathbb{F}}\}.$$

Indeed, take any φ_F from the left-hand side. Since $\Phi_{\mathbb{F}}$ is a cone, $\frac{\varphi_F}{\alpha} \in \Phi_{\mathbb{F}}$. Thus, $\alpha(\frac{\varphi_F}{\alpha}) = \varphi_F \geq \varphi$. By definition, $\varphi_F = \alpha(\frac{\varphi_F}{\alpha})$ belongs to the right-hand side. Conversely, take any $\alpha\varphi_F$ from the right-hand side. Since $\alpha\varphi_F \in \Phi_{\mathbb{F}}$ and $\alpha\varphi_F \geq \varphi$, by definition, $\alpha\varphi_F$ belongs to the left-hand side, as desired.

For all $\varphi \in C(\Delta(\Omega))$ and $\alpha > 0$, the above observation implies that

$$\begin{aligned} V(\alpha\varphi) &= \inf\{V(\varphi_F) |\, \varphi_F \in \Phi_{\mathbb{F}}, \varphi_F \geq \alpha\varphi\} \\ &= \inf\{V(\alpha\varphi_F) |\, \varphi_F \in \Phi_{\mathbb{F}}, \alpha\varphi_F \geq \alpha\varphi\} \\ &= \inf\{\alpha V(\varphi_F) |\, \varphi_F \in \Phi_{\mathbb{F}}, \varphi_F \geq \varphi\} = \alpha V(\varphi). \end{aligned}$$

■

If $W : \Delta(\Delta(\Omega)) \times \mathbb{R} \to \mathbb{R} \cup \{-\infty\}$ is defined as in (11), $\succsim$ is represented by a Costly Subjective Learning Representation.

Lemma 19 *For all* $\pi \in ca_+(\Delta(\Omega))$*,* $W(\pi, t)$ *is homogeneous of degree one in* t*, that is, for all* $\alpha > 0$*,* $W(\pi, \alpha t) = \alpha W(\pi, t)$*.*

Proof. (i) By Lemma 4 and positive homogeneity of V,

$$\begin{aligned} W(\pi, \alpha t) &= \inf_{\varphi \in B(\pi,\alpha t)} V(\varphi) = \inf_{\varphi \in \alpha B(\pi,t)} V(\varphi) = \inf_{\varphi \in B(\pi,t)} V(\alpha\varphi) \\ &= \alpha \inf_{\varphi \in B(\pi,t)} V(\varphi) = \alpha W(\pi, t). \end{aligned}$$

■

For notational convenience, given $\pi \in \Delta(\Delta(\Omega))$, we define a scalar function defined for all $t \in \mathbb{R}$ by $W_\pi(t) = W(\pi, t)$. Since our concern is the infimum, let

$$\operatorname{dom} W_\pi = \{t \in \mathbb{R} \,|\, W_\pi(t) > -\infty\}.$$

Define Π as in (12).

Lemma 20 *For all $\pi \in \Pi$, either $\operatorname{dom} W_\pi = \mathbb{R}$ or $\operatorname{dom} W_\pi = \mathbb{R}_+$.*

Proof. Take any $\pi \in \Pi$. Assume that there exists $t^* < 0$ such that $W(\pi, t^*) > -\infty$. By Lemma 7, $W(\pi, t) \geq W(\pi, t^*) > -\infty$ for all $t \geq t^*$. On the other hand, for all $t < t^* < 0$, by Lemma 19,

$$W(\pi, t) = W(\pi, \frac{t}{t^*}t^*) = \frac{t}{t^*}W(\pi, t^*) > -\infty.$$

Hence, $\operatorname{dom} W_\pi = \mathbb{R}$.

Next, assume that there exists no $t < 0$ such that $W(\pi, t) > -\infty$. That is, $W(\pi, t) = -\infty$ for all $t < 0$. Since $\pi \in \Pi$, there exists $t^* \geq 0$ such that $W(\pi, t^*) > -\infty$. If $t^* = 0$, by Lemma 7, $W(\pi, t) \geq W(\pi, 0) > -\infty$ for all $t \geq 0$. Thus, $\operatorname{dom} W_\pi = \mathbb{R}_+$. If $t^* > 0$, by Lemma 19, for all $t > 0$,

$$W(\pi, t) = W(\pi, \frac{t}{t^*}t^*) = \frac{t}{t^*}W(\pi, t^*) > -\infty.$$

Moreover, since $W(\pi, t)$ is upper semi-continuous in t,

$$W(\pi, 0) = \lim_{t \searrow 0} W(\pi, t) = \lim_{t \searrow 0} \frac{t}{t^*}W(\pi, t^*) = 0 > -\infty.$$

Hence, $\operatorname{dom} W_\pi = \mathbb{R}_+$. ■

By Lemmas 19 and 20, $W(\pi, t)$ can be rewritten as follows: for all $\pi \in \Pi$ and $t > 0$,

$$W(\pi, t) = W(\pi, t \cdot 1) = tW(\pi, 1) = \beta(\pi) \cdot t,$$

where $\beta(\pi) := W(\pi, 1)$. Similarly, for all $\pi \in \Pi$ with $\operatorname{dom} W_\pi = \mathbb{R}$ and $t < 0$,

$$W(\pi, t) = W(\pi, (-t) \cdot -1) = -tW(\pi, -1) = \gamma(\pi) \cdot t,$$

where $\gamma(\pi) := -W(\pi, -1)$. If $\operatorname{dom} W_\pi = \mathbb{R}_+$, define $\gamma(\pi) = \infty$. Thus, for $t \neq 0$, $W(\pi, t)$ is written as

$$W(\pi, t) = \beta(\pi)(t)^+ - \gamma(\pi)(t)^-, \tag{15}$$

where for $t \in \mathbb{R}$, $(t)^+ = \max\{0, t\}$ and $(t)^- = \max\{0, -t\}$ and $-\infty \times 0 = 0$ with convention.

In particular, since $W(\delta_{\overline{p}}, t) = t$ by Lemma 10, $\operatorname{dom} W_{\delta_{\overline{p}}} = \mathbb{R}$, which implies

$$\beta(\delta_{\overline{p}}) = \gamma(\delta_{\overline{p}}) = 1. \tag{16}$$

Lemma 21 *For all* $\pi \in \Pi$, $\beta(\pi)$ *is real-valued, and* $\beta(\pi) \geq 0$*. If* $\text{dom}W_\pi = \mathbb{R}$, $\gamma(\pi)$ *is real valued, and* $\gamma(\pi) \geq 0$*.*

Proof. Since $W(\pi, t)$ is real-valued for any $\pi \in \Pi$ and $t > 0$, $\beta(\pi)$ is real-valued. Next, we show that for all $\pi \in \Pi$, $\beta(\pi) \geq 0$. Suppose contrary that there exists $\pi \in \Pi$ such that $\beta(\pi) = W(\pi, 1) < 0$. For $t > 1$, $W(\pi, t) = W(\pi, 1) \cdot t < W(\pi, 1)$. This contradicts the fact that $W(\pi, t)$ is nondecreasing in t, shown in Lemma 7.

Similarly, since $W(\pi, t)$ is real-valued for any $\pi \in \Pi$ with $\text{dom}\, W_\pi = \mathbb{R}$, $\gamma(\pi)$ is real-valued. Finally, we show that for all such π, $\gamma(\pi) \geq 0$. Since $\langle \varphi_{\{x_0\}}, \pi \rangle = \langle \mathbf{0}, \pi \rangle \geq -1$, we have $W(\pi, -1) \leq 0$. Hence, $\gamma(\pi) = -W(\pi, -1) \geq 0$. ■

The following lemma provides the case of $t = 0$.

Lemma 22 *For any* $\pi \in \Pi$, $W(\pi, 0) = 0$*.*

Proof. By Lemma 21, $W(\pi, t) = \beta(\pi)t$ for all $t > 0$ and $\pi \in \Pi$. Since $W(\pi, t)$ is upper semi-continuous in t,

$$W(\pi, 0) = \lim_{t \searrow 0} W(\pi, t) = \lim_{t \searrow 0} \beta(\pi)t = 0.$$

■

Lemma 23 Π *is closed and convex.*

Proof. To show that Π is closed, let $\pi^n \to \pi$ with $\pi^n \in \Pi$. By Lemma 22, $W(\pi^n, 0) \geq 0$. Since $W : \Delta(\Delta(\Omega)) \times \mathbb{R} \to \mathbb{R}$ is upper semi-continuous, $W(\pi, 0) \geq 0$, which implies $W(\pi, 0) > -\infty$ at π. Hence, $\pi \in \Pi$, as desired.

To show that Π is convex, take $\pi_1, \pi_2 \in \Pi$ and $\alpha \in [0, 1]$. There exist t_i, $i = 1, 2$, such that $W(\pi_i, t_i) > -\infty$. Since W is quasi-concave in (π, t),

$$W(\alpha\pi_1 + (1-\alpha)\pi_2, \alpha t_1 + (1-\alpha)t_2) \geq \min[W(\pi_1, t_1), W(\pi_2, t_2)] > -\infty.$$

Thus, $\alpha\pi_1 + (1-\alpha)\pi_2 \in \Pi$. ■

By Lemma 22, (15) holds for all $t \in \mathbb{R}$. Now, we obtain that

$$V(\varphi) = \max_{\pi \in \Pi} [\beta(\pi)(\langle \varphi, \pi \rangle)^+ - \gamma(\pi)(\langle \varphi, \pi \rangle)^-]. \tag{17}$$

Note that $V(\varphi) \geq 0$ is equivalent to $\langle \varphi, \pi \rangle \geq 0$ for some $\pi \in \Pi$, and $V(\varphi) < 0$ is equivalent to $\langle \varphi, \pi \rangle < 0$ for all $\pi \in \Pi$.

It follows from (17) that $\succsim$ is represented by

$$U(F) = V(\varphi_F) = \max_{\pi \in \Pi} [\beta(\pi)(\langle \varphi_F, \pi \rangle)^+ - \gamma(\pi)(\langle \varphi_F, \pi \rangle)^-]. \tag{18}$$

To obtain a more explicit form of β and γ, we prepare the following lemma.

Lemma 24 *For all* $\pi \in \Pi$*,*

$$\beta(\pi) = \inf_{\{F \in \mathbb{F} \,|\, b_F^u(\pi) > 0\}} \frac{u(x_F)}{\langle \varphi_F, \pi \rangle}.$$

Proof. Since $W(\pi, \alpha)$ is homogeneous in α, for all $\pi \in \Pi$ and $\alpha > 0$, $W(\pi, \alpha) = \alpha W(\pi, 1) = \alpha\beta(\pi)$. By Lemma 9, for any $\pi \in \Pi$ and $\alpha \in \mathbb{R}$

$$W(\pi, \alpha) = \inf_{\varphi \in B(\pi,\alpha)} V(\varphi) = \inf_{\varphi_F \in B(\pi,\alpha)} V(\varphi_F) = \inf_{\{F | \langle \varphi_F, \pi \rangle \geq \alpha\}} u(x_F).$$

We will claim that for any $\pi \in \Pi$ and $\alpha \in \mathbb{R}$,

$$\inf_{\{F | \langle \varphi_F, \pi \rangle \geq \alpha\}} u(x_F) = \inf_{\{F | \langle \varphi_F, \pi \rangle = \alpha\}} u(x_F). \tag{19}$$

Take any F with $\langle \varphi_F, \pi \rangle > \alpha$. For any $\lambda \in (0, 1)$, let $\lambda F + (1 - \lambda)\{x_0\}$ be denoted by λF. Since $\langle \varphi_{\lambda F}, \pi \rangle = \lambda \langle \varphi_F, \pi \rangle$, for any λ sufficiently close to one, $\langle \varphi_F, \pi \rangle > \langle \varphi_{\lambda F}, \pi \rangle > \alpha$. Moreover, since V is positively homogeneous,

$$u(x_F) = V(\varphi_F) > \lambda V(\varphi_F) = V(\lambda \varphi_F) = V(\varphi_{\lambda F}) = u(x_{\lambda F}).$$

That is, if F satisfies $\langle \varphi_F, \pi \rangle > \alpha$, $u(x_F)$ is not a lower bound of $\{u(x_F) \,|\, \langle \varphi_F, \pi \rangle \geq \alpha\}$. Thus, (19) holds.

By the above observations,

$$\begin{aligned}\inf_{\{F \in \mathbb{F} \,|\, b_F^u(\pi) > 0\}} \frac{u(x_F)}{\langle \varphi_F, \pi \rangle} &= \inf_{\alpha > 0} \left(\inf_{\{F | \langle \varphi_F, \pi \rangle = \alpha\}} \frac{u(x_F)}{\alpha} \right) = \inf_{\alpha > 0} \frac{1}{\alpha} \left(\inf_{\{F | \langle \varphi_F, \pi \rangle = \alpha\}} u(x_F) \right) \\ &= \inf_{\alpha > 0} \frac{1}{\alpha} W(\pi, \alpha) = \beta(\pi),\end{aligned}$$

as desired. ■

Lemma 25 *For all* $\pi \in \Pi$ *with* $\operatorname{dom} W_\pi = \mathbb{R}$*,*

$$\gamma(\pi) = \sup_{\{F \in \mathbb{F} \,|\, b_F^u(\pi) < 0\}} \frac{u(x_F)}{\langle \varphi_F, \pi \rangle} = \sup_{\{F \in \mathbb{F} \,|\, u(x_F) < 0\}} \frac{u(x_F)}{\langle \varphi_F, \pi \rangle}.$$

Proof. Note that for all $\pi \in \Pi$ with $\operatorname{dom} W_\pi = \mathbb{R}$ and $\alpha < 0$, $W(\pi, \alpha) = -\alpha W(\pi, -1) = -\alpha\gamma(\pi)$. As in Lemma 24, we have

$$\begin{aligned}\sup_{\{F \in \mathbb{F} | \langle \varphi_F, \pi \rangle < 0\}} \frac{u(x_F)}{\langle \varphi_F, \pi \rangle} &= \sup_{\alpha > 0} \left(\sup_{\{F \in \mathbb{F} \,|\, \langle \varphi_F, \pi \rangle = -\alpha\}} \frac{u(x_F)}{-\alpha} \right) = \sup_{\alpha > 0} -\frac{1}{\alpha} \left(\inf_{\{F \in \mathbb{F} \,|\, \langle \varphi_F, \pi \rangle = -\alpha\}} u(x_F) \right) \\ &= \sup_{\alpha > 0} \left(-\frac{1}{\alpha} W(\pi, -\alpha) \right) = -W(\pi, -1) = \gamma(\pi).\end{aligned}$$

From the functional form, $\gamma(\pi) \geq 0$ for any $\pi \in \Pi$ implies that $V(\varphi_F) < 0$ is equivalent to $\langle \varphi_F, \pi \rangle < 0$ for all $\pi \in \Pi$. Hence, $u(x_F) < 0$ is equivalent to $\langle \varphi_F, \pi \rangle < 0$ for all $\pi \in \Pi$. This implies that $\{F \in \mathbb{F} | \langle \varphi_F, \pi \rangle < 0, u(x_F) < 0\} = \{F \in \mathbb{F} | u(x_F) < 0\}$. Moreover, if $\langle \varphi_F, \pi \rangle < 0$, $u(x_F) < 0$, $\langle \varphi_G, \pi \rangle < 0$, and $u(x_G) \geq 0$, then

$$\frac{u(x_F)}{\langle \varphi_F, \pi \rangle} > 0 \geq \frac{u(x_G)}{\langle \varphi_G, \pi \rangle}.$$

Thus, we have that

$$\gamma(\pi) = \sup_{\{F \in \mathbb{F} | \langle \varphi_F, \pi \rangle < 0\}} \frac{u(x_F)}{\langle \varphi_F, \pi \rangle} = \sup_{\{F \in \mathbb{F} | \langle \varphi_F, \pi \rangle < 0, u(x_F) < 0\}} \frac{u(x_F)}{\langle \varphi_F, \pi \rangle} = \sup_{\{F \in \mathbb{F} | u(x_F) < 0\}} \frac{u(x_F)}{\langle \varphi_F, \pi \rangle}.$$

■

From Lemma 12, $\Pi \subset \Pi(\overline{p})$. Let

$$\Pi^{\mathbb{R}} = \{\pi \in \Pi \,|\, \operatorname{dom} W_\pi = \mathbb{R}\}.$$

Note that β is a real-valued function defined on Π, while γ is a real-valued function defined on $\Pi^{\mathbb{R}}$. Note also that $\Pi^{\mathbb{R}} \neq \emptyset$. In fact, if $\Pi^{\mathbb{R}} = \emptyset$, $W(\pi, t) = -\infty$ for all $t < 0$ and $\pi \in \Pi$. But, since V is normalized, for all $t < 0$,

$$t = V(t\mathbf{1}) = \max_{\pi \in \Pi} W(\pi, \langle t\mathbf{1}, \pi \rangle) = \max_{\pi \in \Pi} W(\pi, t) = -\infty,$$

which is a contradiction. Moreover, $\Pi^{\mathbb{R}}$ is convex. Take any $\pi, \pi' \in \Pi^{\mathbb{R}}$ and $\alpha \in [0, 1]$. Since $W(\pi, t)$ is quasi-concave in (π, t),

$$W(\alpha\pi + (1-\alpha)\pi', -1) \geq \min[W(\pi, -1), W(\pi', -1)] > -\infty,$$

which implies $\operatorname{dom} W_{\alpha\pi + (1-\alpha)\pi'} = \mathbb{R}$. Thus, $\alpha\pi + (1-\alpha)\pi' \in \Pi^{\mathbb{R}}$.

Extend $\beta : \Pi \to \mathbb{R}_+$ and $\gamma : \Pi^{\mathbb{R}} \to \mathbb{R}_+$ to $\Pi(\overline{p})$ by

$$\beta^*(\pi) := \inf_{\{F \in \mathbb{F} \,|\, \langle \varphi_F, \pi \rangle > 0\}} \frac{u(x_F)}{\langle \varphi_F, \pi \rangle}$$

and

$$\gamma^*(\pi) := \sup_{\{F \in \mathbb{F} \,|\, \langle \varphi_F, \pi \rangle < 0\}} \frac{u(x_F)}{\langle \varphi_F, \pi \rangle}.$$

By Lemmas 24 and 25, $\beta^* = \beta$ on Π and $\gamma^* = \gamma$ on $\Pi^{\mathbb{R}}$. Moreover, for all $\pi \in \Pi(\overline{p}) \setminus \Pi$, by definition of β^*,$U(F) \geq \beta^*(\pi) b_F^u(\pi)$ for all F with $b_F^u(\pi) > 0$, and for all $\pi \in \Pi(\overline{p}) \setminus \Pi^{\mathbb{R}}$, $U(F) \geq \gamma^*(\pi) b_F^u(\pi)$ for all F with $b_F^u(F) < 0$. Hence, β^* on $\Pi(\overline{p}) \setminus \Pi$ and γ^* on $\Pi(\overline{p}) \setminus \Pi^{\mathbb{R}}$ are in fact irrelevant for the representation. It follows from (18) that the representation $U(F)$ is rewritten as

$$U(F) = V(\varphi_F) = \max_{\pi \in \Pi(\overline{p})} [\beta^*(\pi)(\langle \varphi_F, \pi \rangle)^+ - \gamma^*(\pi)(\langle \varphi_F, \pi \rangle)^-].$$

Finally, we show that β^* and γ^* have the desired properties.

Lemma 26

$$\gamma^*(\pi) = \sup_{\{F\in\mathbb{F}\,|\,u(x_F)<0\}} \frac{u(x_F)}{\langle\varphi_F,\pi\rangle}.$$

Proof. The proof is the same as in Lemma 25. Since $\gamma^*(\pi) \geq 1 > 0$, the representation implies that $V(\varphi_F) < 0$ is equivalent to $\langle\varphi_F,\pi\rangle < 0$ for all $\pi \in \Pi(\overline{p})$. Hence, $u(x_F) < 0$ is equivalent to $\langle\varphi_F,\pi\rangle < 0$ for all $\pi \in \Pi$. This implies that $\{F \in \mathbb{F}|\langle\varphi_F,\pi\rangle < 0, u(x_F) < 0\} = \{F \in \mathbb{F}|u(x_F) < 0\}$. ■

Lemma 27 *β^* is upper semi-continuous and γ^* is lower semi-continuous.*

Proof. For each fixed F with $\langle\varphi_F,\pi\rangle > 0$, $\frac{u(x_F)}{\langle\varphi_F,\pi\rangle}$ is continuous in π. Since the infimum function among upper semi-continuous functions is also upper semi-continuous, β^* is upper semi-continuous in π.

Similarly, since the supremum function among lower semi-continuous functions is also lower semi-continuous, γ^* is lower semi-continuous in π. ■

Lemma 28 *For all $\pi \in \Pi(\overline{p})$, $\beta^*(\pi) \leq 1$ and $\gamma^*(\pi) \geq 1$.*

Proof. Take any lottery $x \in X$ with $u(x) > 0$. Then, $\varphi_{\{x\}}$ satisfies $\langle\varphi_{\{x\}},\pi\rangle = u(x)$ for all π. By definition of β^*,

$$\beta^*(\pi) \leq \frac{u(x)}{\langle\varphi_{\{x\}},\pi\rangle} = 1.$$

Take any lottery $x \in X$ with $u(x) < 0$. By definition of γ^*,

$$\gamma^*(\pi) \geq \frac{u(x)}{\langle\varphi_{\{x\}},\pi\rangle} = 1.$$

■

Lemma 29 $\beta^*(\delta_{\overline{p}}) = \gamma^*(\delta_{\overline{p}}) = 1$

Proof. Since $\beta^*(\delta_{\overline{p}}) = \beta(\delta_{\overline{p}})$ and $\gamma^*(\delta_{\overline{p}}) = \gamma(\delta_{\overline{p}})$, the result follows from (16). ■

Lemma 30 *For all $\pi,\rho \in \Pi(\overline{p})$, $\pi \trianglerighteq \rho \implies \beta^*(\pi) \leq \beta^*(\rho)$ and $\gamma^*(\pi) \geq \gamma^*(\rho)$.*

Proof. If $\pi \trianglerighteq \rho$, $\{F \in \mathbb{F}\,|\,b^u_F(\rho) > 0\} \subseteq \{F \in \mathbb{F}\,|\,b^u_F(\pi) > 0\}$ and $\langle\varphi_F,\pi\rangle \geq \langle\varphi_F,\rho\rangle$. By definition of β^*, we have $\beta^*(\pi) \leq \beta^*(\rho)$.

If $\pi \trianglerighteq \rho$, it follows from definition of γ^* that only the denominator, which is negative, becomes greater. Hence, we obtain $\gamma^*(\pi) \geq \gamma^*(\rho)$. ■

Consequently,

$$\begin{aligned} U(F) = V(\varphi_F) &= \max_{\pi\in\Pi(\overline{p})} [\beta^*(\pi)(\langle\varphi_F,\pi\rangle)^+ - \gamma^*(\pi)(\langle\varphi_F,\pi\rangle)^-] \\ &= \max_{\pi\in\Pi(\overline{p})} [\beta^*(\pi)(b^u_F(\pi))^+ - \gamma^*(\pi)(b^u_F(\pi))^-] \end{aligned}$$

is an optimal waiting representation. Moreover, this is a canonical representation.

F.1 Necessity

Since the optimal waiting representation is a special case of the Costly Subjective Learning Representation, it is enough to check Reference-Point Independence.

Take any $F \in \mathbb{F}$ and $\alpha \in (0,1)$. Since $b^u_{\{x_0\}}(\pi) = u(x_0) = 0$,

$$\begin{aligned}
&U(\alpha F + (1-\alpha)\{x_0\}) \\
&= \max_{\pi\in\Pi}[\beta(\pi)(b^u_{\alpha F+(1-\alpha)\{x_0\}}(\pi))^+ - \gamma(\pi)(b^u_{\alpha F+(1-\alpha)\{x_0\}}(\pi))^-] \\
&= \max_{\pi\in\Pi}[\beta(\pi)(\alpha b^u_F(\pi) + (1-\alpha)b_{\{x_0\}}(\pi))^+ - \gamma(\pi)(\alpha b^u_F(\pi) + (1-\alpha)b_{\{x_0\}}(\pi))^-] \\
&= \max_{\pi\in\Pi}[\beta(\pi)(\alpha b^u_F(\pi))^+ - \gamma(\pi)(\alpha b^u_F(\pi))^-] \\
&= \alpha \max_{\pi\in\Pi}[\beta(\pi)(b^u_F(\pi))^+ - \gamma(\pi)(b^u_F(\pi))^-] = \alpha U(F).
\end{aligned}$$

Take any $F, G \in \mathbb{F}$ and $\alpha \in (0,1)$. By the above observation,

$$\begin{aligned}
&U(\alpha F + (1-\alpha)\{x_0\}) \geq U(\alpha G + (1-\alpha)\{x_0\}) \\
\iff &\alpha U(F) \geq \alpha U(G) \\
\iff &U(F) \geq U(G),
\end{aligned}$$

that is, the preference U represents satisfies Reference-Point Independence.

G Proof of Theorem 6

By assumption, $\succsim$ admits a canonical optimal waiting representation $U(F)$. Now assume in addition that $\succsim$ satisfies Independence of Degenerate Decisions. As mentioned in Section B.1, the construction of U and V is the same as in de Oliveira, Denti, Mihm, and Ozbek [15]. They show that Independence of Degenerate Decisions together with the other axioms implies translation invariance of $V : \Phi_{\mathbb{F}} \to \mathbb{R}$, that is, $V(\varphi_F + \alpha) = V(\varphi_F) + \alpha$ for all $\alpha \in \mathbb{R}$. Thus, $U(F) = V(\varphi_F) = \max_{\pi\in\Pi}[\beta(\pi)(\langle\varphi_F, \pi\rangle)^+ - \gamma(\pi)(\langle\varphi_F, \pi\rangle)^-]$ satisfies translation invariance. The extension V on $C(\Delta(\Omega))$ satisfies translation invariance because for any $\alpha \in \mathbb{R}$,

$$\begin{aligned}
V(\varphi + \alpha) &= \inf\{V(\varphi_F + \alpha) \,|\, \varphi_F + \alpha \geq \varphi + \alpha\} = \inf\{V(\varphi_F) + \alpha \,|\, \varphi_F \geq \varphi\} \\
&= \inf\{V(\varphi_F) \,|\, \varphi_F \geq \varphi\} + \alpha = V(\varphi) + \alpha.
\end{aligned}$$

It is well-know that if V on $C(\Delta(\Omega))$ is monotone and translation invariant, it is Lipschitz continuous. This implies that V on $C(\Delta(\Omega))$ is uniformly continuous. A slight modification of the proof of Cerreia-Vioglio, Maccheroni, Marinacci, and Montrucchio [6, Theorem 5] shows the following:

Claim 8 *For all $\pi \in \Delta(\Delta(\Omega))$, $\mathrm{domW}_\pi \in \{\emptyset, \mathbb{R}\}$, and $\{W_\pi \,|\, \pi \in \Delta(\Delta(\Omega)), \mathrm{domW}_\pi = \mathbb{R}\}$ is nonempty family of real valued uniformly equicontinuous functions.*[7]

[7] For every $\varepsilon > 0$, there exists $\delta > 0$ such that $|t - t'| < \delta$ implies $|W(\pi, t) - W(\pi, t')| \leq \varepsilon$ for all $t, t' \in \mathbb{R}$ and all $\pi \in \Delta(\Delta(\Omega))$ such that $\mathrm{dom}W_\pi = \mathbb{R}$.

Proof. Since V is uniformly continuous, for all $\varepsilon > 0$, there exists $\delta > 0$ such that if $\|\varphi - \varphi'\| \leq \delta$, $|V(\varphi) - V(\varphi')| \leq \varepsilon$. If $\psi \in C(\Delta(\Omega))$ with $\|\psi\| = 1$, we have

$$V(\varphi + \delta\psi) \leq V(\varphi) + \varepsilon \text{ and } V(\varphi - \delta\psi) \geq V(\varphi) - \varepsilon \tag{20}$$

for all $\varphi \in C(\Delta(\Omega))$ and $\|\psi\| = 1$. Fix $\pi \in \Delta(\Delta(\Omega))$. Since $\pi \in \Delta(\Delta(\Omega))$, there exists $\psi \in C(\Delta(\Omega))$ such that $\|\psi\| = 1$ and $\langle \psi, \pi \rangle \geq 1/2$.[8]

Given $\varphi \in C(\Delta(\Omega))$, W attains its maximum. Hence, we have that $V(\varphi) = W(\pi_\varphi, \langle \varphi, \pi_\varphi \rangle) \in \mathbb{R}$ for some $\pi_\varphi \in \Delta(\Delta(\Omega))$. Given $\varepsilon > 0$, take δ satisfying (20). Let $t \in \mathrm{domW}_\pi$ and $t' \in \mathbb{R}$ with $|t - t'| \leq 1/2$. We have two cases. First, assume that $t' \geq t$. We have

$$\begin{aligned} W_\pi(t) + \varepsilon &= \inf_{\varphi \in B(\pi,t)} V(\varphi) + \varepsilon \geq \inf_{\{\varphi \,|\, \langle \varphi, \pi \rangle \leq t\}} V(\varphi + \delta\psi) = \inf_{\{\varphi \,|\, \langle \varphi - \delta\psi, \pi \rangle \leq t\}} V(\varphi) \\ &= \inf_{\{\varphi \,|\, \langle \varphi, \pi \rangle \leq t + \delta\langle \psi, \pi \rangle\}} V(\varphi) = W_\pi(t + \delta\langle \psi, \pi \rangle) \geq W_\pi(t + \delta/2) \geq G_\pi(t') \geq W_\pi(t), \end{aligned}$$

where the first inequality is by (20), the second inequality follows from $\langle \psi, \pi \rangle \geq 1/2$ and Lemma 7, and the rest follows Lemma 7. Hence, we have $|W_\pi(t) - W_\pi(t')| \leq \varepsilon$.

Second, assume that $t' \leq t$. We have

$$\begin{aligned} W_\pi(t) \geq & W_\pi(t') \geq W_\pi(t - \delta/2) \geq W_\pi(t - \delta\langle \psi, \pi \rangle) = \inf_{\{\varphi \,|\, \langle \varphi, \pi \rangle \leq t - \delta\langle \pi, \psi \rangle\}} V(\varphi) \\ &= \inf_{\{\varphi \,|\, \langle \varphi + \delta\psi, \pi \rangle \leq t\}} V(\varphi) = \inf_{\{\varphi \,|\, \langle \varphi, \pi \rangle \leq t\}} V(\varphi - \delta\psi) \geq \inf_{\varphi \in B(\pi,t)} V(\varphi) - \varepsilon = W_\pi(t) - \varepsilon. \end{aligned}$$

This implies that $|W_\pi(t) - W_\pi(t')| \leq \varepsilon$.

Now, we show the rest statements in the claim. Take $\varepsilon > 0$ and $\pi \in \Delta(\Delta(\Omega))$ such that $\mathrm{domW}_\pi \neq \emptyset$. If $t \in \mathrm{domW}_\pi$, we have $[t - \delta/2, t + \delta/2] \subset \mathrm{domW}_\pi$. Hence, $\mathrm{domW}_\pi = \mathbb{R}$. This implies that $\mathrm{domW}_\pi \in \{\emptyset, \mathbb{R}\}$.

Since given $\varphi \in C(\Delta(\Omega))$, W attains its maximum, we have that $V(\varphi) = W(\pi_\varphi, \langle \varphi, \pi_\varphi \rangle) \in \mathbb{R}$ for some $\pi_\varphi \in \Delta(\Delta(\Omega))$. Thus, we have $\mathrm{domW}_{\pi_\varphi} = \mathbb{R}$. This means that $\{W_\pi \,|\, \pi \in \Delta(\Delta(\Omega)), \mathrm{domW}_\pi = \mathbb{R}\}$ is not empty. Moreover, W_π does not take ∞ for all $\pi \in \Delta(\Delta(\Omega))$ because V is real valued. ■

By Claim 8, we have that $\Pi = \{\pi \in \Delta(\Delta(\Omega)) \,|\, W(t, \pi) > -\infty \text{ for some } t\} = \{\pi \in \Delta(\Delta(\Omega)) \,|\, \mathrm{dom}W_\pi = \mathbb{R}\}$. This implies that $\beta(\pi) = W(1, \pi) \in \mathbb{R}$ and $\gamma(\pi) = -W(-1, \pi) \in \mathbb{R}$ for any $\pi \in \Pi$. Moreover, by Lemma 23, Π is closed and convex.

By the property of β, $\beta(\delta_{\overline{p}}) = 1$ and $\beta(\pi) \in [0, 1]$ for any $\pi \in \Pi$. We show below that if $\beta(\pi) < 1$ for some $\pi \in \Pi$, then $\beta(\pi) < 0$, that is, for all $\pi \in \Pi$, $\beta(\pi) = 1$. Assume that $\beta(\pi) = \inf_F \frac{u(x_F)}{\langle \varphi_F, \pi \rangle} < 1$. There exists F such that $\langle \varphi_F, \pi \rangle > 0$ and $\frac{u(x_F)}{\langle \varphi_F, \pi \rangle} < 1$. Since $\Phi_{\mathbb{F}} + \mathbb{R} = \Phi_{\mathbb{F}}$, for any $\varepsilon > 0$, there exists a menu G such that $\varphi_G = \varphi_F - u(x_F) - \varepsilon$. Since $\langle \varphi_F, \pi \rangle > u(x_F)$, we have that for small enough $\varepsilon > 0$,

$$\langle \varphi_G, \pi \rangle = \langle \varphi_F, \pi \rangle - u(x_F) - \varepsilon > 0.$$

[8]For example, one can take $\psi = \mathbf{1}$.

Moreover, by translation invariance,

$$u(x_G) = V(\varphi_G) = V(\varphi_F - u(x_F) - \varepsilon) = V(\varphi_F) - u(x_F) - \varepsilon < 0.$$

Hence, we have

$$0 > \frac{u(x_G)}{\langle \varphi_G, \pi \rangle} \geq \inf_{\{F' \,|\, \langle \varphi_{F'}, \pi \rangle > 0\}} \frac{u(x_{F'})}{\langle \varphi_{F'}, \pi \rangle} = \beta(\pi).$$

Hence, $\beta(\pi) = 1$ for all $\pi \in \Pi$.

Turn to the premium function. Note that $\gamma(\delta_{\overline{p}}) = 1$ and $\gamma(\pi) \in [1, \infty)$ for any $\pi \in \Pi$. We show that if $\gamma(\pi) > 1$ for some $\pi \in \Pi$, then $\gamma(\pi) = \infty$, that is, for all $\pi \in \Pi$, $\gamma(\pi) = 1$. Assume that $\gamma(\pi) = \sup_F \frac{u(x_F)}{b_F^u(\pi)} > 1$. There exists F such that $u(x_F) < 0$ and $\frac{u(x_F)}{\langle \varphi_F, \pi \rangle} > 1$. Since $u(x_F) < 0$ is equivalent to $\langle \varphi_F, \pi \rangle < 0$ for all $\pi \in \Pi$, $\frac{u(x_F)}{\langle \varphi_F, \pi \rangle} > 1$ implies $\langle \varphi_F, \pi \rangle > u(x_F)$. Since $\Phi_{\mathbb{F}} + \mathbb{R} = \Phi_{\mathbb{F}}$, for any $\varepsilon > 0$, there exists a menu G such that $\varphi_G = \varphi_F - \langle \varphi_F, \pi \rangle - \varepsilon$. Then, we have

$$\langle \varphi_G, \pi \rangle = \langle \varphi_F, \pi \rangle - \langle \varphi_F, \pi \rangle - \varepsilon = -\varepsilon < 0.$$

Moreover, by translation invariance,

$$u(x_G) = V(\varphi_G) = V(\varphi_F) - \langle \varphi_F, \pi \rangle - \varepsilon = u(x_F) - \langle \varphi_F, \pi \rangle - \varepsilon < 0.$$

Hence, we have

$$\frac{u(x_F) - \langle \varphi_F, \pi \rangle - \varepsilon}{-\varepsilon} = \frac{u(x_G)}{\langle \varphi_G, \pi \rangle} \leq \sup_{\{F' \,|\, u(x_{F'}) < 0\}} \frac{u(x_{F'})}{\langle \varphi_{F'}, \pi \rangle} = \gamma(\pi).$$

By $\varepsilon \to 0$, $0 > \langle \varphi_F, \pi \rangle > u(x_F)$ implies that $\gamma(\pi) = \infty$.

As shown above, this optimal waiting representation is reduced to a constrained information representation such that $\beta(\pi) = 1$ and $\gamma(\pi) = 1$ for all $\pi \in \Pi$, that is, U is rewritten as

$$U(F) = \max_{\pi \in \Pi} b_F^u(\pi).$$

H Proof of Corollary 4

Since Independence implies Singleton Independence, Aversion to Contingent Planning, Reference-Point Independence, and Independence of Degenerate Decisions, by Theorem 6, $\succsim$ is represented by the constrained information model,

$$U(F) = \max_{\pi \in \Pi} b_F^u(\pi).$$

Lemma 31 *U is mixture linear.*

Proof. Take any menu F and G and $\alpha \in (0,1)$. Let $\{x_F\}$ be a lottery equivalent of F, that is, $\{x_F\} \sim F$. Since $F \sim \{x_F\}$ and $G \sim \{x_G\}$, Independence implies that $\alpha F + (1-\alpha)G \sim \alpha\{x_F\} + (1-\alpha)\{x_G\}$. Thus,

$$\begin{aligned} U(\alpha F + (1-\alpha)G) &= U(\{\alpha x_F + (1-\alpha)x_G\}) = \alpha u(x_F) + (1-\alpha)u(x_G) \\ &= \alpha U(F) + (1-\alpha)U(G), \end{aligned}$$

as desired. ■

We adopt the same proof of de Oliveira, Denti, Mihm, and Ozbek [15, Corollary 2]. For all menus F, let $\mathcal{C}(F) \subset \Pi$ denote the set of maximizer of $\max_{\pi\in\Pi} b^u_F(\pi)$. Take any F, G, $\alpha \in (0,1)$. Take any $\pi \in \mathcal{C}(\alpha F + (1-\alpha)G)$. By definition of π and mixture linearity,

$$\alpha U(F) + (1-\alpha)U(G) = U(\alpha F + (1-\alpha)G) = b^u_{\alpha F+(1-\alpha)G}(\pi) = \alpha b^u_F(\pi) + (1-\alpha)b^u_G(\pi).$$

Moreover, by the representation, $U(F) \geq b^u_F(\pi)$ and $U(G) \geq b^u_G(\pi)$. By combining $\alpha U(F) + (1-\alpha)U(G) = \alpha b^u_F(\pi) + (1-\alpha)b^u_G(\pi)$ with $U(F) \geq b^u_F(\pi)$, we have

$$\alpha b^u_F(\pi) + (1-\alpha)b^u_G(\pi) \geq \alpha b^u_F(\pi) + (1-\alpha)U(G),$$

which implies $b^u_G(\pi) \geq U(G)$. Together with $U(G) \geq b^u_G(\pi)$, we conclude $U(G) = b^u_G(\pi)$. Thus, $\pi \in \mathcal{C}(G)$. By the symmetric argument, $\pi \in \mathcal{C}(F)$. Therefore, $\mathcal{C}(\alpha F + (1-\alpha)G) \subset \mathcal{C}(F) \cap \mathcal{C}(G)$.

By repeating the same argument finitely many times, we can show the following: For all menus F_i, $i = 1, \cdots, n$, and $\alpha_i > 0$ with $\sum_i \alpha_i = 1$,

$$\mathcal{C}\left(\sum_{i=1}^n \alpha_i F_i\right) \subset \bigcap_{i=1}^n \mathcal{C}(F_i).$$

In particular, we know that $\cap_{i=1}^n \mathcal{C}(F_i) \neq \emptyset$. Since Π is compact, by the finite intersection property of compact sets, it is guaranteed that

$$\bigcap_{F\in\mathbb{F}} \mathcal{C}(F) \neq \emptyset.$$

Take any π^* which belongs to this intersection. Then, $U(F) = b^u_F(\pi^*)$ holds for all F. Consequently, $(u, \{\pi^*\})$ is a fixed information representation for $\succsim$.

References

[1] Aoyama, T. (2020), "Response time and subjective filtration, "working paper.

[2] Anscombe, F. J. and R. J. Aumann (1963), "A definition of subjective probability, "*Annals of Mathematical Statistics*, 34, 199-205.

[3] Barbu, V. and T. Precupanu (2010), *Convexity and Optimization in Banach Spaces*, Fourth Edition, Springer.

[4] Caplin, A., and M. Dean (2015), "Revealed preference, rational inattention, and costly information acquisition, "*American Economic Review*, 105, 2183-2203.

[5] Caplin, A., M. Dean, and J. Leahy (2019), "Rationally inattentive behavior: Characterizing and generalizing Shannon entropy, "working paper.

[6] Cerreia-Vioglio, S., F. Maccheroni, M. Marinacci, and L. Montrucchio (2008), "Complete Monotone Quasiconcave Duality, "working paper.

[7] Cerreia-Vioglio, S., F. Maccheroni, M. Marinacci, and L. Montrucchio (2011), "Complete Monotone Quasiconcave Duality, "*Mathematics of Operations Research*, 36, 321-339.

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

[9] Chambers, C. P., C. Liu, and J. Rehbeck (2020), "Costly information acquisition, "*Journal of Economic Theory*, 186.

[10] Chateauneuf, A., J. H. Faro (2009), "Ambiguity through confidence functions, "*Journal of Mathematical Economics*, 45, 535-558.

[11] Cukierman, A. (1980), "The effects of uncertainty on investment under risk neutrality with endogenous information, "*Journal of Political Economy*, 88, 462-475.

[12] Dekel, E., B. Lipman, and A. Rustichini. Representing preferences with a unique subjective state space. *Econometrica*, Vol. 69, 891-934, 2001.

[13] Denti, T. (2018), "Posterior-separable cost of information, "working paper.

[14] de Oliveira, H. (2014), "Axiomatic foundations for entropic cost of attention, "working paper.

[15] de Oliveira, H., T. Denti, M. Mihm, and K. Ozbek (2017), "Rationally inattentive preferences and hidden information costs, "*Theoretical Economics*, 12, 621-654.

[16] Dillenberger, D., R. V. Krishna, and P. Sadowski (2020), "Subjective information choice processes, "working paper.

[17] Dillenberger, D., R. V. Krishna, and P. Sadowski (2020), "Information-dependent utilities and beliefs, "working paper.

[18] Dillenberger, D., J. S. Lleras, P. Sadowski, and N. Takeoka (2014), "A theory of subjective learning, "*Journal of Economic Theory*, 153, 287-312.

[19] Duraj, J., and Y.-H. Lin (2019), "Costly information and random choice, "working paper.

[20] Ellis, A. (2018), "Foundations for optimal inattention, "*Journal of Economic Theory*, 173, 56-94.

[21] Ergin, H., and T. Sarver (2010), "A unique costly contemplation representation, "*Econometrica*, 78, 1285-1339.

[22] Hyogo, K. (2007), "A subjective model of experimentation, "*Journal of Economic Theory*, 133, 316-330.

[23] Kamenica, E., and M. Gentzkow, "Bayesian Persuasion, "*American Economic Review*, 101, 2590-2615.

[24] Kendall, C. (2018), "The time cost of information in financial markets, "*Journal of Economic Theory*, 176, 118-157.

[25] Lin, Y.-H. (2018), "Stochastic choice and rational inattention, "working paper.

[26] Lu, J. (2016), "Random choice and private information, "*Econometrica*, 84, 1983-2027.

[27] Maccheroni, F., M. Marinacci, and A. Rustichini (2006) "Ambiguity aversion, robustness, and the variational representation of preferences, "*Econometrica*, 74, 1447-1498.

[28] Mihm, M., and K. Ozbek (2018) "Mood-driven choices and self-regulation, "*Journal of Economic Theory*, 176, 727-760.

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

[30] Strazalecki, T. (2013) "Temporal resolution of uncertainty and recursive models of ambiguity aversion, "*Econometrica*, 81, 1039-1074.